

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-041276

(43)Date of publication of application : 08.02.2002

---

(51)Int.Cl. G06F 3/16

G06F 3/00

G10L 15/00

G10L 15/22

H04N 5/00

H04N 5/44

H04N 5/445

H04N 7/025

H04N 7/03

H04N 7/035

H04Q 9/00

-----  
(21)Application number : 2000-222895 (71)Applicant : SONY CORP

(22)Date of filing : 24.07.2000 (72)Inventor : NASHIDA TATSUSHI

BIZEN NAOTO

-----  
(54) INTERACTIVE OPERATION-SUPPORTING SYSTEM, INTERACTIVE  
OPERATION-SUPPORTING METHOD AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To support inputting a user command to an electric household appliance such as a television set and an information appliance.

SOLUTION: According to an operation-supporting system, by making the animation of the character called 'a personified assistant', which shows reaction by a voice synthesis or an animation, a user interface, a friendly feeling is given

to a user, and at the same time the correspondence to complicated commands and the entrance to service can be provided. As the instruction system which is sensuously close to natural language is provided, the user can easily operate the appliance feeling the same sense as his or her usual conversation.

---

LEGAL STATUS [Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

## \* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.

2. \*\*\*\* shows the word which can not be translated.

3. In the drawings, any words are not translated.

---

## CLAIMS

---

[Claim(s)]

[Claim 1] It is the interactive actuation support system which supports the input of the user command to a device, and has a display, the voice input section, the voice output section, and an operation control section. Said operation control section The assistant control means who the assistant who personified is generated [ assistant ] and makes said display appear, The output voice control means which determines the language about which this assistant speaks, synthesizes voice, and carries out an external output through said voice output section, The input speech recognition means which carries out speech recognition of a user's voice inputted through said voice input section, The dialogue management tool which manages the dialogue between both according to a user's words recognized by language and said input speech recognition



means of the assistant determined by said output voice control means, The interactive actuation support system characterized by providing a command interpretation means to specify an intention of a user or an input command based on the contents of a dialogue which said dialogue management tool pursued.

[Claim 2] Said assistant control means is an interactive actuation support system according to claim 1 characterized by determining an assistant's appropriate animation based on the user input command specified by the contents of a dialogue managed by said dialogue management tool, and/or said command interpretation means.

[Claim 3] Said output voice control means is an interactive actuation support system according to claim 1 characterized by determining the language about which this assistant speaks based on the user input command specified by the contents of a dialogue managed by said dialogue management tool, and/or said command interpretation means.

[Claim 4] Said output voice control means is an interactive actuation support system according to claim 1 characterized by determining the language about which this assistant that draws an intention of a user speaks when an intention of a user or an input command cannot be specified in said command interpretation means.

[Claim 5] It is the interactive actuation support system according to claim 1 which is equipped with the connecting means which furthermore connects external instruments, such as a television set and a videocassette recorder, and is characterized by said command interpretation means interpreting the command for functional control of external instruments, such as a channel selection of a program, and an image transcription/playback in a videocassette recorder.

[Claim 6] It is the interactive actuation support system according to claim 1 which is equipped with the means of communications for furthermore connecting with communication media, such as an external network and the general telephone line, and is characterized by said input speech recognition means performing said recognition processing of the voice data received via communication media.

[Claim 7] It is the interactive actuation support system according to claim 1 which is further equipped with the means of communications for connecting with communication media, such as an external network and the general telephone line, and this e-mail exchange means to exchange electronic mails via communication media, and is characterized by said output voice control means determining the language about which an assistant speaks based on the contents of reception mail.

[Claim 8] It is the interactive actuation support system according to claim 1 which said dialogue management tool manages a certain user's input voice as a

message to other users, and is characterized by said output voice control means determining the language about which an assistant speaks based on this message.

[Claim 9] Said assistant control means is an interactive actuation support system according to claim 1 characterized by stationing the personified assistant in the room where various kinds of objects including the link to an information resource are scattered.

[Claim 10] It is the interactive actuation support system according to claim 1 which said assistant control means stations the personified assistant in the room where various kinds of objects including the link to an information resource are scattered, answers that the user has had the interest to an archive medium including the link to the music content put on this room, and is characterized by said command interpretation means interpreting the input of a command as playback of this music content.

[Claim 11] It is the interactive actuation support system according to claim 1 which answers that said command interpretation means interpreted the command input from a user, and is characterized by said assistant control means making an assistant appear in said display.

[Claim 12] It is the interactive actuation support system according to claim 1 which is equipped with the connecting means which furthermore connects a

television set, answers that said command interpretation means interpreted the channel channel selection from a user as an input command, and is characterized by said assistant control means making the assistant who carried the viewing window of the tuned-in program appear.

[Claim 13] It is the interactive actuation support system according to claim 1 which is equipped with the connecting means which furthermore connects a television set, answers that said command interpretation means interpreted the channel change from a user as an input command, and is characterized by said assistant control means arranging each viewing window of a switchable program around an assistant approximate circle annular.

[Claim 14] Furthermore, the interactive actuation support system according to claim 13 characterized by making the program viewing window which follows the channel directions by the user, is made to rotate each viewing window on this approximate circle ring, and answers and corresponds to channel channel selection decision zoom in.

[Claim 15] It is the interactive actuation support system according to claim 1 which is equipped with the connecting means which connects the secondary storage which furthermore accumulates and reproduces program contents, answers that said command interpretation means interpreted playback of the image transcription program from a user as an input command, and is

characterized by said assistant control means making the assistant carrying the binder which indicated the program recorded on videotape by the list appear.

[Claim 16] Furthermore, the interactive actuation support system according to claim 15 characterized by making the image transcription program viewing window which answers and corresponds to selection decision of an image transcription program to reproduce zoom in.

[Claim 17] It is the interactive actuation support system according to claim 1 which is equipped with the connecting means which furthermore connects a television set, answers that said command interpretation means interpreted the channel change from a user as an input command, and is characterized by said assistant control means making the assistant carrying the program chart which has arranged the switchable program in the shape of a matrix appear.

[Claim 18] Furthermore, the interactive actuation support system according to claim 17 characterized by making the program viewing window which answers and corresponds to channel channel selection decision zoom in.

[Claim 19] The interactive actuation support system according to claim 17 characterized by generating a matrix-like program chart using the electronic program guide (EPG) distributed as a part of data broadcasting.

[Claim 20] It is the interactive actuation support system according to claim 1 which is further equipped with the connecting means which connects a television

set, the means of communications for connecting with communication media, such as an external network and the general telephone line, and this e-mail exchange means to exchange electronic mails via communication media, and is characterized by for said assistant control means answering e-mail reception, and making an e-mail reception display appear on said display.

[Claim 21] Furthermore, the interactive actuation support system according to claim 1 characterized by having a transliteration means to change the kanji into a kana about the text data displayed on the screen of said display.

[Claim 22] The interactive actuation support system according to claim 1 characterized by having further the means of communications for connecting with communication media, such as an external network and the general telephone line, and an authentication means to perform authentication processing of this information terminal connected via communication media.

[Claim 23] The interactive actuation support system according to claim 1 characterized by having further the connecting means which connects a television set, and an extract means to extract text out of the program contents which made it receive.

[Claim 24] The interactive actuation support system according to claim 23 characterized by displaying the text extracted by said extract means on the display program of other programs.

[Claim 25] It is the interactive actuation exchange approach which supports the input of the user command to a device or other devices by which external connection was made concerned on the device equipped with a display, the voice input section, and the voice output section. The assistant control step who the assistant who personified is generated [ assistant ] and makes said display appear, The output voice-control step which determines the language about which this assistant speaks, synthesizes voice, and carries out an external output through said voice output section, The input speech recognition step which carries out speech recognition of a user's voice inputted through said voice input section, The dialogue management step which manages the dialogue between both according to a user's words recognized by language and said input speech recognition step of the assistant determined by said output voice-control step, The interactive actuation exchange approach characterized by providing the command interpretation step which specifies an intention of a user or an input command based on the contents of a dialogue pursued by said dialogue management step.

[Claim 26] The interactive actuation exchange approach according to claim 25 characterized by determining an assistant's appropriate animation based on the user input command specified at said assistant control step by the contents of a dialogue managed by said dialogue management step, and/or said command

interpretation step.

[Claim 27] The interactive actuation exchange approach according to claim 25 characterized by determining the language about which this assistant speaks based on the user input command specified at said output voice-control step by the contents of a dialogue managed by said dialogue management step, and/or said command interpretation step.

[Claim 28] The interactive actuation exchange approach according to claim 25 characterized by determining the language about which this assistant that draws an intention of a user speaks at said output voice-control step when an intention of a user or an input command cannot be specified in said command interpretation step.

[Claim 29] Said device is the interactive actuation exchange approach according to claim 25 characterized by having the connecting means which connects external instruments, such as a television set and a videocassette recorder, further, and interpreting the command for functional control of external instruments, such as a channel selection of a program, and an image transcription/playback in a videocassette recorder, at said command interpretation step.

[Claim 30] Said device is the interactive actuation exchange approach according to claim 25 which is equipped with the means of communications for connecting



with communication media, such as an external network and the general telephone line, further, and is characterized by performing recognition processing of said voice data received via communication media at said input speech recognition step.

[Claim 31] Said device is the interactive actuation exchange approach according to claim 25 characterized by determine the language with which it has further the means of communications for connect with communication media , such as an external network and the general telephone line , and this e-mail exchange means to exchange electronic mails via communication media , and an assistant speaks about them based on the contents of reception mail at said output voice control step .

[Claim 32] The interactive actuation exchange approach according to claim 25 which manages the input voice of the user who is said dialogue management step as a message to other users, and is characterized by determining the language about which an assistant speaks based on this message at said output voice-control step.

[Claim 33] The interactive actuation exchange approach according to claim 25 characterized by stationing the personified assistant at said assistant control step in the room where various kinds of objects including the link to an information resource are scattered.

[Claim 34] The interactive actuation exchange approach according to claim 25 characterized by to station the personified assistant at said assistant control step in the room where various kinds of objects including the link to an information resource are scattered, to answer that the user has had the interest to an archive medium including the link to the music content put on this room, and to interpret the input of a command as playback of this music content at said command interpretation step.

[Claim 35] It is the interactive actuation exchange approach according to claim 25 which answers having interpreted the command input from a user in said command interpretation step, and is characterized by said assistant control step making an assistant appear in said display.

[Claim 36] It is the interactive actuation exchange approach according to claim 25 which said device is equipped with the connecting means which connects a television set further, answers having interpreted the channel channel selection from a user as an input command in said command interpretation step, and is characterized by said assistant control step making the assistant who carried the viewing window of the tuned-in program appear.

[Claim 37] Said device is the interactive actuation exchange approach according to claim 25 which is equipped with the connecting means which connects a television set further, answers having interpreted the channel change from a

user as an input command in said command interpretation step, and is characterized by said assistant control step arranging each viewing window of a switchable program around an assistant approximate circle annular.

[Claim 38] Furthermore, the interactive actuation exchange approach according to claim 37 characterized by making the program viewing window which follows the channel directions by the user, is made to rotate each viewing window on this approximate circle ring, and answers and corresponds to channel channel selection decision zoom in.

[Claim 39] It is the interactive actuation exchange approach according to claim 25 which said device is equipped with the connecting means which connects the secondary storage which accumulates and reproduces program contents further , answers having interpreted playback of the image transcription program from a user as an input command by said command interpretation step , and is characterized by for said assistant control step to make the assistant carrying the binder which indicated the program recorded on videotape by the list appear .

[Claim 40] Furthermore, the interactive actuation exchange approach according to claim 39 characterized by making the image transcription program viewing window which answers and corresponds to selection decision of an image transcription program to reproduce zoom in.

[Claim 41] Said device is the interactive actuation exchange approach according

to claim 25 which is equipped with the connecting means which connects a television set further, answers having interpreted the channel change from a user as an input command by said command interpretation step, and is characterized by said assistant control step making the assistant carrying the program chart which has arranged the switchable program in the shape of a matrix appear.

[Claim 42] Furthermore, the interactive actuation exchange approach according to claim 41 characterized by making the program viewing window which answers and corresponds to channel channel selection decision zoom in.

[Claim 43] The interactive actuation exchange approach according to claim 41 characterized by generating a matrix-like program chart using the electronic program guide (EPG) distributed as a part of data broadcasting.

[Claim 44] Said device is the interactive actuation exchange approach according to claim 25 characterized by having further the connecting means which connects a television set , the means of communications for connecting with communication media , such as an external network and the general telephone line , and this e-mail exchange means to exchange electronic mails via communication media , answering e-mail reception at said assistant control step , and making an e-mail reception display appear on said display .

[Claim 45] Furthermore, the interactive actuation exchange approach according

to claim 25 characterized by having the transliteration step which changes the kanji into a kana about the text data displayed on the screen of said display.

[Claim 46] The interactive actuation exchange approach according to claim 25 characterized by having further a communication link step linked to communication media, such as an external network and the general telephone line, and the authentication step which performs authentication processing of this information terminal connected via communication media.

[Claim 47] Said device is the interactive actuation exchange approach according to claim 25 characterized by having further the extract step which is equipped with the connecting means which connects a television set, and extracts text out of the program contents which made it receive.

[Claim 48] The interactive actuation exchange approach according to claim 47 characterized by displaying the text extracted by said extract step on the display program of other programs.

[Claim 49] On the device equipped with a display, the voice input section, and the voice output section It is the storage which stored physically the computer software described to perform interactive actuation exchange processing which supports the input of the user command to a device or other devices by which external connection was made concerned on computer system in the computer-readable format. The assistant control step who said computer

software generates [ assistant ] the assistant who personified and makes said display appear, The output voice-control step which determines the language about which this assistant speaks, synthesizes voice, and carries out an external output through said voice output section, The input speech recognition step which carries out speech recognition of a user's voice inputted through said voice input section, The dialogue management step which manages the dialogue between both according to a user's words recognized by language and said input speech recognition step of the assistant determined by said output voice-control step, The storage characterized by providing the command interpretation step which specifies an intention of a user or an input command based on the contents of a dialogue pursued by said dialogue management step.

---

## DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the actuation support system and the actuation exchange approach of supporting the user command input to

household-electric-appliances devices and information machines and equipment, such as a television set, and relates to the interactive actuation support system and the interactive actuation exchange approach of performing user command input to various kinds of connection devices to a dialogue especially.

[0002] Furthermore, in detail, this invention relates to the interactive actuation support system and the interactive actuation exchange approach which were made to input the user command to a device in a natural format through the assistant who personified, and relates to the interactive actuation support system and the interactive actuation exchange approach the dialogue of the voice input base with the assistant who personified especially performs command input.

[0003]

[Description of the Prior Art] Conventionally, the thing of the type which a user operates by a hand was in use like the control panel, and a keyboard and a mouse, the command input, i.e., the user interface, to information machines and equipment including various household-electric-appliances devices, such as a television set, or a personal computer. Furthermore, recently, in an advance of cognitive engineering, such as improvement in the operation input of a processor and speech recognition, the interactive user interface of the voice input base is also spreading conjointly.

[0004] In the case of the user interface by the former hand actuation, since the direct and most important command can be inputted to a device, actuation is trustworthy, but a user's burden with which must understand the actuation technique of a device to some extent, and must become skilled, and a user is burdened is excessive.

[0005] For example, in television or other AV equipments, "the interface by the fingertip" which controls the menu by the cross-joint key etc. is a core. With such a user input format, it is expected in network household electric appliances that it becomes complicated to take correspondence.

[0006] Although the user interface which used the commander has permeated in current, if a ground wave, a satellite, the Internet, HAVI (Home Audio/Video Interoperability: common command system of a digital AV equipment), etc. tend to perform many channels and many control, its number of switches will increase too much, and they will become actuation top difficulty. Moreover, by combining with a multifunctional switch and a menu screen, although it becomes possible to reduce the number of switches to some extent, actuation will become complicated rather.

[0007] Moreover, since a command can be specified by analyzing a demand of a user based on the recognition result of input voice to a device side in the case of the user interface by voice input, the burden to the user at the time of device



actuation is mitigated. However, a user has to talk toward a microphone in the scene in which talks and a partner is not, and cannot call it human being's natural action, but may be accompanied by mental pain on the occasion of a dialogue input.

[0008] So, recently, the assistant who personified is made to appear on a display screen, and the interactive actuation support system of realizing command input to a device in the format of talking by a user taking a stand against the assistant on a screen is appearing.

[0009] For example, especially therefore to JP,11-65814,A, it is indicating about the user interface which was rich in presence and a sense of reality by detecting the sound and the direction of a sound source which a user emits, and controlling the image of a virtual living thing according to the detection result (for example, the direction which a virtual living thing follows a sound source and voice always carries out being gazed at).

[0010] Moreover, it is indicated about the agent equipment which provides JP,11-37766,A with the personified agent with a communication function with the operator within a car. According to this agent equipment, the action which the agent personified from the study effectiveness based on the hysteresis of not only the situation of the present car and operator but the past doubled with the situation can be done, and a car can aim at communication with an operator now.

[0011] It is possible for high-level interactive processing to become possible by strengthening of the arithmetic proficiency of a calculating machine etc., and to give an intelligence also to the assistant on a screen recently. For example, it is possible it not only to operate the simple input command which can be constituted from one word like channel selection, initiation of an image transcription and playback, or a halt, but to follow the context of the contents of conversation with a user, and to perform complicated actuation over two or more phases.

[0012] However, a system by which an assistant presents such an advance situation of actuation on a screen is not yet developed, but a user merely cannot but gaze at a screen and cannot but wait for the response from a system. I will be, even when the processing which requires the response time is ordered, and it has misunderstanding with failure of a device.

[0013] In order for a user to operate a device based on a dialogue with an assistant, it is desirable that the command structure of sensibility near natural language can be easily inputted on actuation.

[0014]

[Problem(s) to be Solved by the Invention] The purpose of this invention is to offer the outstanding actuation support system and the outstanding actuation exchange approach of supporting the user command input to

household-electric-appliances devices and information machines and equipment, such as a television set.

[0015] The further purpose of this invention is to offer the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach of performing user command input to a device to a dialogue.

[0016] The further purpose of this invention is to offer the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach of inputting the user command to a device in a natural format through the assistant who personified.

[0017] The further purpose of this invention is to offer the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach the dialogue of the voice input base with the assistant who personified can perform command input.

[0018] The further purpose of this invention is to offer the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach of feeding back the advance situation of command actuation that the user inputted by the dialogue of the voice input base with an assistant to a user.

[0019]

[Means for Solving the Problem] This invention is made in consideration of the above-mentioned technical problem. The 1st side face It is the interactive

actuation support system which supports the input of the user command to a device, and has a display, the voice input section, the voice output section, and an operation control section. Said operation control section The assistant control means who the assistant who personified is generated [ assistant ] and makes said display appear, The output voice control means which determines the language about which this assistant speaks, synthesizes voice, and carries out an external output through said voice output section, The input speech recognition means which carries out speech recognition of a user's voice inputted through said voice input section, The dialogue management tool which manages the dialogue between both according to a user's words recognized by language and said input speech recognition means of the assistant determined by said output voice control means, It is the interactive actuation support system characterized by providing a command interpretation means to specify an intention of a user or an input command based on the contents of a dialogue which said dialogue management tool pursued.

[0020] You may make it said assistant control means determine an assistant's appropriate animation here based on the user input command specified by the contents of a dialogue managed by said dialogue management tool, and/or said command interpretation means.

[0021] Moreover, you may make it said output voice control means determine

the language about which this assistant speaks based on the user input command specified by the contents of a dialogue managed by said dialogue management tool, and/or said command interpretation means.

[0022] Moreover, when an intention of a user or an input command cannot be specified in said command interpretation means, you may make it said output voice control means determine the language about which this assistant that draws an intention of a user speaks.

[0023] Moreover, the interactive actuation support system may be equipped with the connecting means which connects external instruments, such as a television set and a videocassette recorder, further. In such a case, you may make it said command interpretation means also interpret the command for functional control of external instruments, such as a channel selection of a program, and an image transcription/playback in a videocassette recorder.

[0024] Moreover, the interactive actuation support system may be equipped with the means of communications for connecting with communication media, such as an external network and the general telephone line, further. In such a case, said input speech recognition means may be made to perform recognition processing of said voice data received via communication media.

[0025] Moreover, the interactive actuation support system may be further equipped with the means of communications for connecting with communication

media, such as an external network and the general telephone line, and this e-mail exchange means to exchange electronic mails via communication media. In such a case, you may make it said output voice control means determine the language about which an assistant speaks based on the contents of reception mail.

[0026] Moreover, you may make it said dialogue management tool manage a certain user's input voice as a message to other users. In such a case, you may make it said output voice control means determine the language about which an assistant speaks based on this message.

[0027] Moreover, you may make it, as for said assistant control means, station the personified assistant in the room (character room) where various kinds of included objects are scattered in the link to an information resource. For example, it answers that the user has had the interest to an archive medium including the link to the music content put on this room, and you may make it said command interpretation means interpret the input of a command as playback of this music content.

[0028] Moreover, it answers that said command interpretation means interpreted the command input from a user, and you may make it said assistant control means make an assistant appear in said display.

[0029] Moreover, the interactive actuation support system may be equipped with

the connecting means which connects a television set further. In such a case, it answers that said command interpretation means interpreted the channel channel selection from a user as an input command, and you may make it said assistant control means make the assistant who carried the viewing window of the tuned-in program appear.

[0030] Or it answers that said command interpretation means interpreted the channel change from a user as an input command, and you may make it said assistant control means arrange each viewing window of a switchable program around an assistant approximate circle annular. Furthermore, the channel directions by the user are followed, each viewing window is rotated on this approximate circle ring, and you may make it make the program viewing window which answers and corresponds to channel channel selection decision zoom in.

[0031] Moreover, the interactive actuation support system may be equipped with the connecting means which connects the secondary storage which accumulates and reproduces program contents further. The media storage with which a videocassette recorder, a hard disk, DVD-RAM drive CD-R / W drive, etc. can record mass media contents is contained in the secondary storage said here. In such a case, it answers that said command interpretation means interpreted playback of the image transcription program from a user as an input command, and you may make it said assistant control means make the assistant carrying

the binder which indicated the program recorded on videotape by the list appear.

Furthermore, you may make it make the image transcription program viewing window which answers and corresponds to selection decision of an image transcription program to reproduce zoom in.

[0032] Moreover, the interactive actuation support system may be equipped with the connecting means which connects a television set further. In such a case, it answers that said command interpretation means interpreted the channel change from a user as an input command, and you may make it said assistant control means make the assistant carrying the program chart which has arranged the switchable program in the shape of a matrix appear. Furthermore, you may make it make the program viewing window which answers and corresponds to channel channel selection decision zoom in. Moreover, you may make it generate a matrix-like program chart using the electronic program guide (EPG) distributed as a part of data broadcasting.

[0033] Moreover, the interactive actuation support system may be further equipped with the connecting means which connects a television set, the means of communications for connecting with communication media, such as an external network and the general telephone line, and this e-mail exchange means to exchange electronic mails via communication media. In such a case, said assistant control means answers e-mail reception, and you may make it



make an e-mail reception display appear on said display.

[0034] Moreover, you may have further a transliteration means to change the kanji into a kana about the text data displayed on the screen of said display.

[0035] Moreover, the interactive actuation support system may be further equipped with the means of communications for connecting with communication media, such as an external network and the general telephone line, and an authentication means to perform authentication processing of this information terminal connected via communication media.

[0036] Moreover, the interactive actuation support system may be further equipped with the connecting means which connects a television set, and an extract means to extract text out of the program contents which made it receive. In such a case, you may make it display the text extracted by said extract means on the display program of other programs.

[0037] The 2nd side face of this invention moreover, on the device equipped with a display, the voice input section, and the voice output section The assistant control step who it is the interactive actuation exchange approach which supports the input of the user command to a device or other devices by which external connection was made concerned, and the assistant who personified is generated [ assistant ] and makes said display appear, The output voice-control step which determines the language about which this assistant speaks,

synthesizes voice, and carries out an external output through said voice output section, The input speech recognition step which carries out speech recognition of a user's voice inputted through said voice input section, The dialogue management step which manages the dialogue between both according to a user's words recognized by language and said input speech recognition step of the assistant determined by said output voice-control step, It is the interactive actuation exchange approach characterized by providing the command interpretation step which specifies an intention of a user or an input command based on the contents of a dialogue pursued by said dialogue management step.

[0038] You may make it determine an assistant's appropriate animation at said assistant control step here based on the user input command specified by the contents of a dialogue managed by said dialogue management step, and/or said command interpretation step.

[0039] Moreover, you may make it determine the language about which this assistant speaks at said output voice-control step based on the user input command specified by the contents of a dialogue managed by said dialogue management step, and/or said command interpretation step.

[0040] Moreover, when an intention of a user or an input command cannot be specified in said command interpretation step, you may make it determine the language about which this assistant that draws an intention of a user speaks at

said output voice-control step.

[0041] Moreover, when it has the connecting means to which said device connects external instruments, such as a television set and a videocassette recorder, further, you may make it also interpret the command for functional control of external instruments, such as a channel selection of a program, and an image transcription/playback in a videocassette recorder, at said command interpretation step.

[0042] Moreover, when it has means of communications for said device to connect with communication media, such as an external network and the general telephone line, further, at said input speech recognition step, it may be made to perform recognition processing of said voice data received via communication media.

[0043] Moreover, when said device is further equipped with the means of communications for connecting with communication media, such as an external network and the general telephone line, and this e-mail exchange means to exchange electronic mails via communication media, you may make it determine the language about which an assistant speaks based on the contents of reception mail at said output voice-control step.

[0044] Moreover, the input voice of the user who is said dialogue management step is managed as a message to other users, and you may make it determine

the language about which an assistant speaks based on this message at said output voice-control step.

[0045] Moreover, you may make it station the personified assistant at said assistant control step in the room where various kinds of objects including the link to an information resource are scattered. For example, it answers that the user has had the interest to an archive medium including the link to the music content put on this room, and you may make it interpret the input of a command as playback of this music content at said command interpretation step.

[0046] Moreover, it answers having interpreted the command input from a user in said command interpretation step, and you may make it said assistant control step make an assistant appear in said display.

[0047] Moreover, when it has the connecting means to which said device connects a television set further, it answers having interpreted the channel channel selection from a user as an input command in said command interpretation step, and you may make it said assistant control step make the assistant who carried the viewing window of the tuned-in program appear.

[0048] Or it answers having interpreted the channel change from a user as an input command in said command interpretation step, and you may make it said assistant control step arrange each viewing window of a switchable program around an assistant approximate circle annular. Furthermore, the channel

directions by the user are followed, each viewing window is rotated on this approximate circle ring, and you may make it make the program viewing window which answers and corresponds to channel channel selection decision zoom in.

[0049] Moreover, when it has the connecting means to which said device connects secondary storage, such as a videocassette recorder which accumulates and reproduces program contents further, it answers having interpreted playback of the image transcription program from a user as an input command by said command interpretation step, and you may make it said assistant control step make the assistant carrying the binder which indicated the program recorded on videotape by the list appear. Furthermore, you may make it make the image transcription program viewing window which answers and corresponds to selection decision of an image transcription program to reproduce zoom in. A hard disk and the media storage with which DVD-RAM drive CD-R / W drive can record mass media contents are contained in the secondary storage said here in addition to a videocassette recorder.

[0050] Moreover, it answers that said device interpreted the channel change from a user as an input command by said command interpretation step when it had the connecting means which connects a television set further, and you may make it said assistant control step make the assistant carrying the program chart which has arranged the switchable program in the shape of a matrix appear.

Furthermore, you may make it make the program viewing window which answers and corresponds to channel channel selection decision zoom in. Moreover, you may make it generate a matrix-like program chart using the electronic program guide (EPG) distributed as a part of data broadcasting.

[0051] Moreover, when said device is further equipped with the connecting means which connects a television set, the means of communications for connecting with communication media, such as an external network and the general telephone line, and this e-mail exchange means to exchange electronic mails via communication media, e-mail reception is answered and you may make it make an e-mail reception display appear on said display at said assistant control step.

[0052] Moreover, the interactive actuation exchange approach may be further equipped with the transliteration step which changes the kanji into a kana about the text data displayed on the screen of said display.

[0053] Moreover, the interactive actuation exchange approach may be further equipped with the communication link step linked to communication media, such as an external network and the general telephone line, and the authentication step which performs authentication processing of this information terminal connected via communication media.

[0054] Moreover, when it has the connecting means to which said device

connects a television set, the interactive actuation exchange approach may be further equipped with the extract step which extracts text out of the program contents which made it receive. Moreover, you may make it display the text extracted by said extract step on the display program of other programs.

[0055] The 3rd side face of this invention moreover, on the device equipped with a display, the voice input section, and the voice output section It is the storage which stored physically the computer software described to perform interactive actuation exchange processing which supports the input of the user command to a device or other devices by which external connection was made concerned on computer system in the computer-readable format. The assistant control step who said computer software generates [ assistant ] the assistant who personified and makes said display appear, The output voice-control step which determines the language about which this assistant speaks, synthesizes voice, and carries out an external output through said voice output section, The input speech recognition step which carries out speech recognition of a user's voice inputted through said voice input section, The dialogue management step which manages the dialogue between both according to a user's words recognized by language and said input speech recognition step of the assistant determined by said output voice-control step, It is the storage characterized by providing the command interpretation step which specifies an intention of a user or an input

command based on the contents of a dialogue pursued by said dialogue management step.

[0056] The storage concerning the 3rd side face of this invention is a medium which offers computer software physically in a computer-readable format to the computer system of the versatility which can perform various program codes, for example. Attachment and detachment of CD (Compact Disc), FD (Floppy Disc), MO (Magneto-Optical disc), etc., etc. are free for such a medium, and it is a storage of portability. Or it is also technically possible to provide specific computer system with computer software in a computer-readable format via transmission media, such as a network (for a network not to ask distinction of wireless and a cable), etc.

[0057] Such a storage defines the collaboration-relation on the structure of the computer software and the storage for realizing the function of computer software predetermined in a computer system top, or a function. If it puts in another way, by installing predetermined computer software in computer system through the storage concerning the 3rd side face of this invention, on computer system, a collaboration-operation is demonstrated and the same operation effectiveness as the interactive actuation support system or approach concerning each 1st [ of this invention ] or 2nd side face can be acquired.

[0058]



[Function] According to the actuation support system and approach concerning this invention, by making animation of the character called the personified assistant who performs reaction by speech synthesis or animation into a user interface, while giving familiarity to a user, the entry to the correspondence and service to a complicated instruction can be offered. Moreover, since it has the command structure of sensibility near natural language, a user can operate a device easily with the same feeling as the usual conversation.

[0059] The purpose, the description, and advantage of further others of this invention will become [ rather than ] clear by detailed explanation based on the example and the drawing to attach of this invention mentioned later.

[0060]

[Embodiment of the Invention] Hereafter, the example of this invention is explained in detail, referring to a drawing.

[0061] The hardware configuration of the interactive actuation support system 1 with which operation of this invention is presented is typically shown in drawing 1 . This system is constituted as a receiving set for television sets like a set top box (STB), and is connected to the television set. The interactive actuation support system 1 can perform a dialogue with a user through the "assistant" who mentions later, and can interpret an explicit or potential intention of a user based on this dialogue, and can offer exchange service of user actuation of a channel

selection, an image transcription, playback, etc., etc. Hereafter, each part is explained, referring to drawing 1 .

[0062] The CC section 11 is an operation control unit which controls the actuation in the interactive actuation support system 1 in generalization according to a predetermined control program, for example, performs management of a dialogue with the manifestation of an assistant's behavior pattern, and the user and assistant who used I/O of voice and an image as the base etc. in the generation list of the assistant as a user's partner.

[0063] The function shown below is included in the control program performed in the CC section 11. That is, the input command obtained by the voice input and speech recognition through the (1) microphone 22 is followed, and it is the motion control of each part in a system 1.

(2) Character control about the control (4) assistant of the control (3) tuner 15 of the various external instruments connected through the input/output interface 17 (generation of the animation corresponding to the input command obtained by speech recognition)

(5) Speech synthesis (conversion to the sound signal of the voice data which a character utters.) However, it is made for the animation and voice of a character, such as a motion of opening, to synchronize.

(6) Control of the screen output through the control (9) monitor 25 of the voice

output through the control (8) loudspeaker 21 through the communication link interface 18 of the data for data broadcasting of Control (7) (Electric Programming Guide) EPG and others of connection with an external network etc.

(10) Conversion of text data based on the processing (12) user profile of the text data used on the control (11) electronic mail by the input command through remote control (not shown), EPG, and a wide area network (it is conversion to [kanji] -> [a hiragana] for example, to for children)

(13) Image measurement (the score progress on sports programs, such as baseball and soccer, and extract of the time-of-day data from a screen display) based on the data accompanied by a video signal, various services based on image recognition (a notice, time-of-day doubling, etc. of score progress)

(14) Basic setup of the synthetic (16) system of the texture chosen in the bit map conversion (15) texture database which used the font chosen in the font database based on text data, and the bit map of a font (the brightness of a screen, sound volume, various I/O, etc.)

[0064] A tuner 15 performs with the directions from the CC section 11, tuning, i.e., a channel selection, of the broadcast wave of a predetermined channel. The received broadcast wave is divided into a part for a part for image data division, and sound data division. The screen output of the image data is carried out through the image-processing section 16 at a monitor 25. Moreover, sound

output of the sound data is carried out to a loudspeaker 21 through the speech synthesis section 14 (or you may be line-out).

[0065] The recording apparatus 12 is keeping data required for an assistant's image generation or generation of a behavior pattern. The following information is accumulated in storage data, for example. Namely, advertising contents which should be inserted into the character and the feeling based on the hysteresis and hysteresis of the conversation of the past of the user's user profile information (4) user and assistant who converse with the information (3) assistant of a layout or others' about the character room in which a data (2) assistant required in order to generate (1) assistant's 3D character image information and its animation works, or other exchanges, a study data (5) assistant, or a character room [0066] Moreover, a store 12 also performs storage of various kinds of databases (not shown), such as a font database and a texture database, besides the information about an assistant. A font database manages the various fonts used in EPG, a notice plate, an electronic mail, etc. Moreover, as for a texture database, various kinds of textures used in EPG, a notice plate, an electronic mail, etc. are managed (bit map data of 2D etc.).

[0067] Although only a single assistant may be made to appear on a system 1, a different assistant for every user may be prepared. Two or more characters from which young and old of both sexes, character, etc. are different beforehand are

prepared, and it relates with a user profile etc. by the own alternative, and you may make it make automatic selection according to a user profile, or register at the time of an initial log in. Moreover, an assistant's study and historical data are also prepared for every user, and it may be the same assistant or you may make it take another reaction for every user.

[0068] The speech recognition section 13 recognizes a user's voice for the voice input signal through audio input units, such as a microphone 22, as text information, and it performs further analysis processing of the user input changed into text format, using a language database (not shown). A text is classified per word by morphological analysis etc., and in syntax and semantic discourse analysis, he acquires language information, such as syntactic information and discourse information, understands a user input command etc., and, more specifically, it outputs to the CC section 11.

[0069] An input/output interface 17 is equipment for connecting external instruments, such as a videocassette recorder 23 and a personal computer (PC) 24, to the actuation support system 1. For example, according to interface specifications, such as i-link (IEEE1394), external connection of one or more AV equipments or the information machines and equipment can be made.

[0070] The communication link interface 18 is equipment for interconnecting a system 1 with other host computer systems on an external network. An external

network is a wide area network like the Internet. On a network, the WWW server which offers the WWW (World Wide Web) resource described in the HTML format, the advertising servers which offer the mail server which offers e-mail transceiver service to each user account, and the advertising contents updated every moment are scattered. In this example, at least one server on a network is a character server which carries out the De Dis tribute of the character data, such as an image about the assistant who offers interactive actuation exchange, and animation, character, a behavioral model, the charge or for nothing.

[0071] In addition, on a network, the commerce server, the Internet Service Provider, etc. which serve settlement of accounts (electronic banking) for the information offer server which carries out the offer service of the information, such as Timely which offers information changed every moment, such as Daily and the stock quotations which offer high information on sexual, such as Season Database which collected a public engine's etc. services, Weekly which offers a weather report and program guidance per week, news, and an advertisement, per Japanese instancy, and a money order, traffic information, and a PD selling list exist.

[0072] For example, if it is a TCP/IP (Transmission Control Protocol/Internet Protocol) network, since the resource offered in each server will be identified in a URL (Uniform Resource Locator) format, a system 1 can download these

information resources according to predetermined protocols, such as HTTP (Hyper TextTransfer Protocol). Therefore, the actuation support system 1 concerning this example is the timing of periodic or arbitration, can re-download character data in use, and can update an image, character, a function, etc. of an assistant.

[0073] The transaction which the actuation support system 1 can perform via a network is illustrated below.

(1) An image transcription vicarious execution demand in case there is no means (for example, media, such as a video tape) to record on videotape to carry out the download (5) program image transcription of the download (4) texture data of the download (3) font data of the character which constitutes the updating (2) assistant of a control program who drives each part in a system 1 (refer to the JP,2000-162320,A specification already transferred to these people about an image transcription vicarious execution system)

(6) Character control via the acquisition (9) electronic commerce (10) network of the service use (8) weather report of the analysis of a user profile, and a user customize (7) public engine, program guidance, news, traffic information, an advertisement, etc. (voice, an electronic mail, the Web page for control, etc. are used)

[0074] A modem 19 is equipment for transmitting the computer data of a digital

format via general public line circuits, such as PSTN (Public Switched Telephone Network), and performs modulation processing to an analog signal, and recovery processing to a digital signal.

[0075] On the general public line, exchange connection of the installation telephone of for example, each home is carried out. Moreover, as for the general public line, the radiotelephony network interconnects. Therefore, the interactive actuation support system 1 concerning this example can perform the data exchange between an installation telephone or a cellular phone. Moreover, the assistant generated automatically in the CC section 11 can interpret an input command based on a dialogue with the user through a cellular phone.

[0076] The CC section 11 follows the speech recognition result by the speech recognition section 13, controls the actuation in a system 1, and realizes the dialogue between a user and a system 1.

[0077] For example, when the voice by which the user input was carried out is conversation with an assistant, according to the study and historical data, and the behavioral model about an assistant, it opts for an assistant's behavior and opts for the reaction of the assistant by voice or the dynamic image.

[0078] An assistant's voice is compounded by the speech synthesis section 14, and an external output is carried out by the loudspeaker 41. When the sound of the program under broadcasting is flowing, it superimposes on this and you may



make it utter an assistant's voice.

[0079] Moreover, with reference to 3D character information and animation information, image composition of the motion of an assistant is carried out in the image-processing section 14. You may make it switch a background (scene) with reference to character room information depending on the case at this time. Moreover, it may be made to indicate with one or more program images while broadcasting an assistant's image and the character room in which an assistant is by superposition (the after-mentioned is yielded about the detail of this point).

[0080] Moreover, when the voice by which the user input was carried out is a channel switch, the CC section 11 transmits a channel number to a tuner 15, and is used for channel selection actuation.

[0081] Moreover, when the voice by which the user input was carried out is the command (for example, initiation and termination of an image transcription or playback actuation, a rapid traverse, rewinding, search, a file transfer) to external instruments, such as a videocassette recorder 23 and a personal computer 24, the CC section 11 publishes a command to the device which corresponds by input/output interface 17 course.

[0082] Moreover, when the voice by which the user input was carried out directs access to a wide area network, the access request to the host system with which it was specified on the network by the communication link interface 18 course is

transmitted. For example, when networks are TCP/IP networks, such as the Internet, an access request is described in a URL format. In such a case, a user may emit the language (for example, title of a homepage) which could read out URL or was connected with URL and a meaning. It is changed into URL after speech recognition is carried out [ voice / which was inputted through the assistant ] in the case of the latter.

[0083] In addition, although not illustrated, the actuation support system 1 may receive command input with remote control like the conventional AV equipment. In such a case, it has the decoder which interprets the receive section and received data of the wireless (infrared radiation) data sent from remote control, and data processing of data and the command which were decoded is carried out in the CC section 11.

[0084] Although the actuation support system 1 concerning this example performs command input from a user with the voice base, the description is in the point of performing interactive processing as agency, about the "assistant" generated on the screen of a monitor 25 in that case. The assistant who says here is 3D character equipped with the animation function. Moreover, interactive processing shall include interpreting the conversation (every day) of a user and an assistant and extracting a command, giving user feedback through an assistant's reaction, etc.

[0085] According to the actuation support system 1 concerning this example, by making the personified assistant who performs reaction by speech synthesis or animation into a user interface, while giving familiarity to a user, the entry to the correspondence and service to a complicated instruction can be offered. Moreover, since the actuation support system 1 is equipped with the command structure of sensibility near natural language, a user can operate a device easily with the same feeling as the usual conversation.

[0086] In drawing 2 , the command-processing system in the actuation support system 1 concerning this example is illustrated.

[0087] The voice input from the user through a microphone is recognized as text information in the speech recognition section 13, and analysis processing of the user input command changed into text format is performed further, using a language database (not shown).

[0088] "Character control" for generating the behavior of the assistant in response to a user input in this example, for example, "Device control" which performs directions of the external instrument of the videocassette recorder connected to the actuation support system 1, and others of operation, "Speech synthesis" which generates answerback of the assistant in response to a user input, "mail which transmit and receive mail through a network", the "notice plate" which performs message switching among two or more users

(unspecified), etc. are prepared as a command in which a user input is possible.

[0089] Character control controls the device by which external connection was made at the actuation support system 1 or this to the input command of the voice base from a user with an assistant's behavior (in other words considering an assistant's behavior as user feedback).

[0090] In drawing 3 , the character control system on the actuation support system 1 concerning this example is illustrated.

[0091] As shown in this drawing, functional command, such as "a channel channel selection", "a channel change-over", "AV equipment control", "e-mail reading", "e-mail writing", "notice plate reading", "notice plate writing", and "Ambient", is realizable through character control.

[0092] Here, "Ambient" carries out actuation with the suitable character which constitutes an assistant at the time of a waiting state, or means a function which presses for a user input by speech synthesis.

[0093] Moreover, in case the function in which a character reads out text information, such as "e-mail reading" and "notice plate reading", is realized, it is desirable to take the synchronization of voice and animation so that lip sync (form of opening of pronunciation) may be performed automatically.

[0094] The after-mentioned is yielded about the procedure of realizing functional command through character control.

[0095] A character notifies a user of a current condition through means, such as speech synthesis. Moreover, a user is a network course, or when the actuation support system 1 is being operated by remote control through a cellular phone, means, such as an electronic mail, may be made to perform the notice of a condition.

[0096] On the same actuation support system 1, even if it prepares only the single character which constitutes an assistant, the character customized according to the individual for every user may be prepared. The model and animation of each character can be updated through means of communications, such as a network, an archive medium, and broadcast. Moreover, you may make it map other information contents, such as an advertisement, dynamically to the texture of the clothes of each character.

[0097] The principle configuration for performing command processing in the voice base in drawing 4 on the actuation support system 1 concerning this example is illustrated.

[0098] Speech recognition is performed in the speech recognition section 13, and the voice inputted through the microphone 22 is changed into the information on the text base.

[0099] Based on text information, the CC section 11 performs user interaction and understands the command inputted by the user with the gestalt near natural

language.

[0100] And the CC section 11 controls the actuation of an AV equipment by which external connection was made to the actuation support system 1 according to a command interpretation result. Furthermore, the animation of speech synthesis or a character can be generated and user feedback of the activation situation of a directive command etc. can be carried out.

[0101] Here, some examples about the command interpretation processing of the voice input base performed on the actuation support system 1 concerning this example are shown below.

[0102] (Example 1) the voice input from a user -- "-- video 1 -- video 2 -- dubbing -- "

-> the inputted voice -- a text -- changing -- further -- "video" and "1" -- "-- since -- " -- "-- video" and "2" -- "-- alike -- " -- "-- dubbing" -- "-- carrying out -- " -- token division is carried out.

-> This instruction gestalt is classified into two devices and one instruction "dubbing" "video", "1", and "2."

-> this -- analyzing lexical -- "video" and "1" -- "-- since -- " -- "-- video" and "2" -- "-- alike -- " -- "-- the command for appliance control dubbing" is generated.

[0103] (Example 2) Voice input from a user "dub video"

-> Token division of the inputted voice is carried out at "video", "dub", and

"dubbing."

-> since the instruction "dubbing" is included in this instruction gestalt, it is presumed that there are two device devices. Then, further, the voice output said to where is performed and the voice input "the video 2 from video 1" from the user to this is received from where.

-> this -- "video" and "1" -- "-- since -- " -- "-- video" and "2" -- "-- passing -- " -- token division is carried out.

-> The insufficient thing is reinputted to a token interpretation and it is made an instruction called dubbing from video 1 to video 2.

[0104] (Example 3) Register a synonym into a token interpretation and cope with an extensive expression. For example, in the Tokyo area, "one channel" shall cover all the following expressions.

It is \*\*\*\*\* synthesis television \*\*\*\*\*-synthesis \*\*\*\*\*-synthesis television \*\*\*\*\*--- in my \*\*\*\*\* my \*\*\*\*. -- -- [0105] (Example 4) There are a single, double, and triple -- in an instruction gestalt. a single instruction gestalt -- one mere instruction gestalt -- it is -- for example, -- "-- it can obtain and sew and is \*\*\*\*\*-", "being \*\* in my \*\*\*\*\*", etc. Moreover, a double instruction gestalt consists of one device and one instruction gestalt. for example, "television" to a television image transcription and an "image transcription" -- "-- carrying out --." (however -- "-- carrying out -- " -- it is omissible) -- an "image transcription" of as

opposed to [ although recorded on videotape ] playback, and "the thing carried out" -- "-- showing -- " -- etc. -- it is equivalent to a double instruction gestalt. Moreover, a triple instruction gestalt consists of two devices and one instruction gestalt. "video" and "1" which direct dubbing from video 1 to video 2 -- "-- since -- " -- "-- video" and "2" -- "-- alike -- " -- "-- "DVD" and "" which direct dubbing from DVD to dubbing" and video 1, "video", and "1" -- "-- alike -- " -- "-- copy" etc. is equivalent to a triple instruction gestalt. [ for example, ]

[0106] The processing flow of character control is roughly shown in drawing 5 .

[0107] For example, activation of the character which constitutes an assistant is carried out by switching on the power source of a television monitor 25. An assistant stands still for example, in a living room (or his single room), and he stands by until there are user directions of "a change-over of a channel", an "animation image transcription", "what was recorded on videotape being seen (playback)", "e-mail being written", "e-mail being read", "Web being seen (retrieval of WWW information space)", "a message (writing to a notice plate or reading)", "service", etc., etc.

[0108] A user can give directions in a natural language format by the voice input through a microphone 22 as opposed to an assistant. Or giving an assistant directions through a cellular phone from a going-out place etc. is also permitted. However, from a remote environment, since functions, such as channel selection



and playback of an image transcription program, do not make semantics, the function permitted to cellular phones, such as an "animation image transcription", "e-mail being seen", and "service", may be restricted.

[0109] The example of the display screen just behind powering on of a television monitor 25 is shown in drawing 6 . [0110] Immediately after starting, an assistant is "Ambient", i.e., a waiting state, and suitable actuation is carried out or he performs behavior which presses for a user input by speech synthesis.

[0111] the example of illustration -- "-- giving up -- " -- the assistant character called has appeared in an assistant's room (character room). Each assistant has a default or the character by which a user setup was carried out. "-- giving up -- " -- \*\*\*\* -- for example, conditions as shown below are set up.

[0112]

[Equation 1] The fine boy who gives up and lives in -- television later on. It is growing up every day. Most things are controllable, if anything is known if it is naturally the thing of television and it is connected with television, since television is made into the roost. Since it is full of curiosity, various things are always put on his room, and the newest information also tends to be in disorder! Action which a translation sometimes does not understand is also carried out. (since it is a child well) Although there are seemingly a friend and an associate somehow, if there are still few people who were able to meet well -- Think that it

says well that people are good, or is a lovely interface!!

[0113] In the means ax room, various goods (object), such as a magazine and a toy, are scattered. These objects have the link reference to various services, such as an advertisement and sales (distribution of data contents besides physical shopping or music is included). That is, indoors what is accessed is scattered and it has suggested whether if a user says something, things [ what kind of ] can be accessed and it will meet.

[0114] Use of an advertisement or other information resources is attained by preparing an assistant's room. Moreover, an assistant is pulling out a user's profile by the dialogue with a user, and the correspondence to finer service is attained. Moreover, it can respond to a user's expression, a peculiar peculiarity, etc.

[0115] for example, speech recognition and the information resource a command interpretation is carried out and a goods object carries out [ an information resource ] link reference are called by a user's giving up and asking the thing of the goods which are alike and receive and are interested. The dress of an object or an assistant which are indoors scattered may be changed day by day.

[0116] Moreover, when the actuation support system 1 is accumulating the music content which can be performed beforehand, as shown in drawing 6 , CD

and a CD player may be set in the means ax room. In this case, each scattered CD (archive medium) has the link to the corresponding music content. and if a user gives up as "the CD what ?", is alike and receives and asks, it may give up and may answer, "it is O\*\*" (refer to drawing 7 ), and playback of a music content may be started so that it may perform with the CD player in a screen.

[0117] or it judges that the user is holding interest in the musical piece on the hit chart concerned to it having been alike, and having received and asked, and CD of a hit chart is put on the floor of the means ax room, and it may be [ a user gives up as "the CD what ?", and ] made to publish the request of CD purchase (online shopping) etc. to a predetermined server, or to download the music content concerned from a predetermined site.

[0118] Moreover, in the example shown in drawing 6 , the poster of an automobile is stuck on the wall surface of the means ax room. Thus, the poster stuck on the room is link reference to advertising media or advertising media. For example, purchase induction and presentation of advertising information can be performed with the utterance gestalt by the assistant.

[0119] Moreover, in the example shown in drawing 6 , the globe is installed in the means ax room. A metaphor called, the system to offer information, i.e., WWW, developed on a scale of the earth, can be given to a globe. in this case, by giving up, it being alike, receiving and addressing about a globe, a WWW browser can

start (not shown) and WWW retrieval can be performed.

[0120] Moreover, in the example shown in drawing 6 , the television set is installed in the means ax room. This has a metaphor about the actual television monitor 25, and you may make it display the televising program of a last channel (namely, channel set up at the end) (in the example of illustration, the "channel B" as a last channel has projected).

[0121] Moreover, although not shown in drawing 6 , the mail box may be installed in the means ax room. The mail box has a metaphor about the in-tray of an electronic mail, and when reception mail is accumulated, you may make it display mail into a mail box.

[0122] In addition, it does not go to the display screen of such [ directly ] a means ax room by powering on, but the utterance and remote control input by the user are answered, it gives up, and you may make it arrive at the ax room.

[0123] according to the actuation support system 1 concerning this example -- giving up -- \*\* -- by making the assistant who says into a user interface, while giving familiarity to a user, the entry to the correspondence and service to a complicated instruction can be offered.

[0124] Subsequently, various kinds of actuation exchange processings by the actuation support system 1 through the assistant who takes speech synthesis and the reaction by 3D animation are explained.

[0125] (1) the power-source OFF condition (however, the actuation support system 1 under operation) of power-source ON/OFF television of television -- it is -- a user -- "-- giving up -- television -- \*\*\*\*\* -- " -- etc. -- ask an assistant in a natural language format like.

[0126] such a user input -- the speech recognition section 13 -- setting -- speech recognition -- and a command interpretation is carried out and the power source of television is switched on.

[0127] The initial screen just behind powering on may be a means ax room as shown in drawing 6 . Moreover, the method of the means ax appearance on a screen also prepares some variations, and you may make it use a suitable thing alternatively according to every user, a user's situation, the weather, etc. As the method of an appearance of an assistant, the following variations are mentioned, for example.

[0128] - the means ax room -- "yes" and \*\*\*\*\* - \*\*\* -- while saying -- giving up --  
\*\*\*\*\*.

- It gives up and is loitering in \*\*\*\*\* without saying anything.

- "\*\*\*\*" -- "-- obtaining - and" -- it goes into a standby (Ambient) condition joyfully, saying.

[0129] Here, if voice input of the purport a user "wants to watch television further" is performed, speech recognition and a command interpretation will be

made in the speech recognition section 13. And while giving up and performing acknowledge of \*\*, "O.K.", or "RYOKAI" in a voice output format, television imagery projects greatly by the last channel (namely, channel chosen at the time of the last power source OFF). For example, the last channel has projected on the imagination television set installed in the means ax room (above-mentioned), and this imagination television monitor screen is expanded gradually, and you may make it operate so that the actual whole monitor display may be occupied soon.

[0130] on the other hand -- the time of turning OFF the power source of television -- a user -- "-- giving up -- television. \*\* -- carrying out -- " -- etc. -- what is necessary is just to ask an assistant in a natural language format like

[0131] such a user input -- the speech recognition section 13 -- setting -- speech recognition -- and a command interpretation is carried out and the power source of television is intercepted.

[0132] Moreover, he gives up on a screen, he wears \*\*\*\*\* and the face which seems to be somewhat sad, the back is turned, and it walks and goes in the screen depth direction, and may be made to perform production of disappearing from a screen after receiving voice input until it intercepts the power source of television.

[0133] (2) The direct channel selection a user instructs a specific channel to be

clearly, and zapping-channel selection of wanting to see some out of the program under current broadcast are mentioned to selection of the selector channel of a channel.

[0134] A direct channel selection: In a direct channel selection, since a channel to see based on the voice input from a user is specified, it is desirable that the command structure of sensibility near natural language can be easily inputted on actuation.

[0135] this example -- giving up -- \*\* -- by making the assistant who says into a user interface, while giving familiarity to a user, the entry to the correspondence and service to a complicated instruction can be offered. For example, only by the first voice input, when an intention of a user cannot fully be interpreted, a more natural interface can be built by the personification-interaction which asks back the place which is not known.

[0136] The example of direct channel selection actuation is shown below.

[0137] (Example 1) A user "give up and it is the NHK \*\*\*\*\*. One channel"

-> It gives up and he is expression "NHK. 1 game ?"

-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0138] which gives up and disappears (Example 2) A user "give up and it is educational \*\*\*\*\*. Three

channels"

-> It gives up and is expression "education. 3 game ?"

-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0139] which gives

up and disappears (Example 3) A user "give up and it is Japanese tele \*\*\*\*\*.

Four channels"

-> It gives up and is an expression "day call. 4 game ?"

-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0140] which gives

up and disappears (Example 4) A user "give up and it is the TBS \*\*\*\*\*. Six

channels"

-> It gives up and he is expression "TBS. 6 game ?"

-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0141] which gives

up and disappears (Example 5) A user "give up and it is the Fuji Television \*\*\*\*\*.

Eight channels"

-> It gives up and he is expression "FUJI. 8 game ?"



-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0142] which gives up and disappears (Example 6) A user "give up and show in the tele morning. Ten channels"

-> It gives up and is an expression "tele morning. 10 game ?"

-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0143] which gives up and disappears (Example 7) A user "give up and it is the TV Tokyo \*\*\*\*\*. 12 channels"

-> It gives up and he is expression "TV Tokyo. 12 game ?"

-> User "it meets (acknowledge)"

-> (it waits for the next voice input)

-> (predetermined time -- there was what [ no ] -- answering) [0144] which gives up and disappears In drawing 8 - drawing 11 , the example of a screen display when performing direct channel selection actuation through an assistant is illustrated.

[0145] First, the user of the actuation support system 1 presupposes that it was viewing and listening to the baseball relay broadcast program under televising

as a viewer of a TV program (refer to drawing 8 ). thus, under broadcast reception -- a user -- for example, -- "-- giving up -- eight channels -- carrying out -- " -- "-- giving up -- Fuji Television -- carrying out -- " -- or it is saying a channel name changing clearly like "Fuji Television", and a direct channel selection change is made through an assistant.

[0146] In the speech recognition section 13, speech recognition of the voice input from such a user is carried out, and it is interpreted as a command. And where the panel (or window) which displayed the program under present broadcast by means \*\* as an assistant and the directed channel is carried, it appears on monitor display (refer to drawing 9 ). In this condition, the channel of a change-over place is still in the condition of a temporary channel selection.

[0147] and an assistant -- "-- be fastidious -- words, such as "?", -- an acknowledge request -- carrying out -- on the other hand -- from a user side -- "-- so -- " -- etc. -- if acknowledge [ like ] is obtained, the channel of a change-over place will change in the condition of channel selection decision from the condition of a temporary channel selection.

[0148] By answering having carried out channel selection decision, when the assistant extrudes the panel which displays the original channel program (this example baseball relay broadcast) toward space right-hand side, the program under televising appears gradually by the channel (this example eight channels)

of the change-over place hidden in that back (refer to drawing 10 and drawing 11).

[0149] However, the format which removes the program of the original channel is not limited to especially the thing made to slide to a longitudinal direction as shown in drawing 10. for example, the means as an assistant -- it may pull down or be made to pull up the panel of the original program toward a screen lower part.

[0150] On single monitor display, drawing 8 - drawing 11 are expressing the program under broadcast in the original channel list as each channel of a change-over place at coincidence, and are expressing direct channel selection modification intelligibly visually. A multiple decoding function realizes the coincidence display of two or more programs.

[0151] In addition, an effect of advertising is expectable by inserting a LOGO and advertising information in the dress of the assistant who appears for direct channel selection modification etc. Of course, you may make it change an assistant's dress according to the times.

[0152] In the form of the flow chart shows the procedure for realizing the user interface by direct command format to drawing 12. According to this procedure, the actuation support system 1 can realize a more natural user interface by the interaction of asking back the personified assistant as agency about the part

which is not known only by the voice input from a user. Hereafter, it explains according to this flow chart.

[0153] First, in the speech recognition section 13, speech recognition of the voice input from the user through a microphone 22 is carried out, and an input keyword is extracted (step S1).

[0154] Subsequently, the corresponding user command is specified by searching the database prepared for every category (step S2). The database according to category is kept for example, in storage 12, and can update the contents via a network.

[0155] And only by the voice input to current, it is insufficient, and when a user command cannot be specified, it progresses to the branching No of the decision block S3, the voice output which reflects upon the information which runs short through an assistant is performed (step S5), and it returns to step S1, and waits for the next voice input from a user.

[0156] On the other hand, when a user command is able to be specified by the voice input to current, a command is published to the corresponding devices (for example, a television monitor 25, a videocassette recorder 23, etc.) (step S4).

And it returns to step S1 and waits for the next voice input from a user.

[0157] zapping channel selection: -- the case where he wants to perform a zapping-channel selection -- a user -- "-- giving up -- all -- showing -- " -- "--

giving up -- it shines [ something ] 8 -- directions like ?" which do not dare specify a channel for a channel are performed to an assistant.

[0158] Such voice directions are answered and an assistant, i.e., the multi-view which gives up and can command a whole view of each program under \*\*\*\*\* and current broadcast, is displayed on monitor display.

[0159] The example of the multi-view screen which can command a whole view of the program under broadcast by each channel is shown in drawing 13 . In the example shown in this drawing, it is arranged in the shape of a circular ring so that the panel which displays the program under current broadcast by each channel may give up and may enclose \*\* in the means as an assistant. You may make it arrange in a head a last channel (namely, channel tuned in at the end) immediately after changing to a multi-view screen, for example.

[0160] a such multi-view screen top -- a user -- the means as an assistant -- it is alike, and it receives and directions by voice input can be given in a natural language format.

[0161] for example, "12 channel -- showing -- " -- by giving the directions which specified channels, such as "12 channels", the "TV Tokyo \*\*\*\*\*", and "TV Tokyo", sequential migration of each program display panel is carried out on a circular ring so that the program display panel of the corresponding channel may come to the forefront side on monitor display. Moreover, it can move each

program display panel at a time to the front or back on [ one ] a circular ring with directions of a "front", a "degree", etc., etc. (refer to drawing 14 ).

[0162] Signs that it looked at the multi-view screen from the upper part virtually are shown in drawing 15 . As shown in this drawing, each program display panel is arranged by abbreviation regular intervals along the circular ring top which a program display panel moves to a means ax section indoor. Moreover, the program display panel of the channel by which current temporary selection is made is arranged at the method of the forefront of a circular ring.

[0163] Moreover, as shown in this drawing, as for some circular rings, the gap where a program display panel is not arranged is prepared. Although this gap carries out sequential migration along with a circular ring at each program display-panel list according to the channel migration directions from a user, a user can catch having moved the channel visually by following the location which this gap moved by the eye.

[0164] If predetermined time (for example, for several seconds) passes in the condition [ that a certain program display panel has come to the method of the forefront ], the corresponding channel will be in a temporary channel selection condition, and highlighting of the panel will be carried out. Channel C will be in a temporary channel selection condition, and signs that highlighting is carried out are shown in drawing 16 .

[0165] the time of the program display panel of a desired channel coming to the method of the forefront of a screen, as for a user -- "curve" and "be fastidious" -- "-- swerving -- showing -- " -- "-- be fastidious -- showing -- " -- etc. -- the channel selection of the channel of a temporary channel selection condition can be decided by emphasizing like and saying.

[0166] If a channel selection is decided, while tuning of the this chosen channel will be performed, the program display panel of this channel is expanded gradually, and comes to occupy the whole monitor display soon.

[0167] The channel selection of the channel C under temporary channel selection is decided in drawing 17 , and the corresponding program display panel shows signs that it expands gradually to it.

[0168] On the multi-view screen as shown in drawing 13 , zapping channel selection modification is visually expressed intelligibly by expressing the program under present broadcast as each channel on single monitor display at coincidence. A multiple decoding function realizes the coincidence display of two or more programs.

[0169] In addition, the program display panel which appears on a multi-view screen does not necessarily need to be a program under current broadcast, for example, may be the playback program taken out of the image transcription program accumulated on a videocassette recorder or other recording devices.

[0170] (3) The actuation support system 1 concerning image transcription this example has connected one or more sets of videocassette recorders 23 as an external instrument, and can specify them as an image transcription place of the program which received. Moreover, the storage 12 of the interior which consists of hard disk drives etc. can be specified as an image transcription place.

[0171] Moreover, the actuation support system 1 concerning this example can offer a more natural user interface by the interaction of the voice base by considering the personified assistant as agency. Therefore, image transcription directions of a desired program can be performed by asking an assistant in a natural language format, without depending on the conventional device actuation of remote control etc.

[0172] Signs that a program image transcription is directed with the voice base to an assistant are shown in drawing 18.

[0173] As shown in this drawing (a), the user is viewing and listening to the program of for example, a baseball relay broadcast. thus, the time of wanting to have come to record on videotape during program reception -- an assistant -- receiving -- "-- recording on videotape -- " -- "-- this -- recording on videotape -- " -- \*\* -- directions of the voice base to say are emitted.

[0174] Speech recognition of the inputted voice is carried out in the speech recognition section 13, a command interpretation is carried out further, and a



system 1 specifies that an input command is the image transcription demand under program reception.

[0175] And it searches for a free area to a videocassette recorder 23 or an integral hard disk etc. by which external connection was made. After securing an image transcription place, "RYOKAI" etc. is acknowledged through an assistant by speech synthesis.

[0176] Moreover, as shown in drawing 18 (b), on monitor display, the counter in which image transcription time amount is expressed as the image transcription icon which shows that it is under image transcription appears.

[0177] (4) Besides the program under image transcription present reception of a broadcast schedule program, the image transcription of a broadcast schedule program, i.e., timed recording, can also be performed.

[0178] The actuation support system 1 concerning this example can offer a more natural user interface by the interaction of the voice base by considering the personified assistant as agency. Therefore, timed recording can be directed by asking an assistant in a natural language format, without depending on the conventional device actuation of remote control etc.

[0179] The example of a screen display in the case of setting up timed recording through an assistant is shown in drawing 23 and drawing 24 .

[0180] a user -- for example, -- "-- since it gives up and starts -- what and \*\* -- " --

-- giving up -- \*\*\*\* and \*\* -- " -- "give up and show EPG (Electric Programming Guide)" -- -- giving up -- what from 8:00, and \*\* -- " -- like -- voice input of the words which pull out a broadcast schedule is carried out through an assistant.

[0181] In the speech recognition section 13, speech recognition of the voice input from such a user is carried out, and it is interpreted as a command. And the program list of the broadcast schedules constituted means \*\* as an assistant and in the shape of a matrix is displayed (refer to drawing 23 ). A program list is generable using EPG distributed as for example, data for data broadcasting.

[0182] After an assistant shows monitor display a program chart as shown in drawing 23 , he may be made to read this out.

[0183] A user can find out a program to record on videotape easily out of the race card by which it was indicated by the list. and -- for example, -- -- the channel D image transcription from 8:00 -- carrying out -- " -- etc. -- directions of a natural language format [ like ] are stopped as an assistant, and carry out voice input through \*\*.

[0184] In the speech recognition section 13, speech recognition of the voice input from such a user is carried out, and it is interpreted as a command. And highlighting of the selected channel D is carried out. Moreover, the column of means \*\* as an assistant and Channel D is pointed out, and the voice output of the words which stimulate a check like "this ?" is carried out.

[0185] If there is an affirmative response from a user to an acknowledge request, image transcription reservation will be set up. In this case, you may make it not only to carry out highlighting of the column of the program by which image transcription assignment was carried out, but display an image transcription reservation icon (not shown).

[0186] (5) The actuation support system 1 concerning playback this example of an image transcription program can offer a more natural user interface by the interaction of the voice base by considering the personified assistant as agency. Therefore, image transcription directions of a desired program can be performed by asking an assistant in a natural language format, without depending not only for the image transcription of a program but for playback of an image transcription program on the conventional device actuation of remote control etc.

[0187] Moreover, when the image transcription place of a program is the storage with which random access is permitted like hard disk equipment, the image transcription program of arbitration can be taken out and playback can be started.

[0188] In drawing 19 - drawing 22 , the example of a screen display when performing playback actuation of an image transcription program through an assistant is illustrated.

[0189] First, the user of the actuation support system 1 presupposes that it was viewing and listening to the song program under televising as a viewer of a TV

program (refer to drawing 19 ). thus, under broadcast reception -- a user -- "8  
recorded -- showing -- " -- playback directions can be performed through an  
assistant by saying playback of an image transcription program clearly like  
"video \*\*\*\*\* recorded on videotape."

[0190] In the speech recognition section 13, speech recognition of the voice  
input from such a user is carried out, and it is interpreted as a command. And the  
image which put the contents (thumbnail of a program) of means \*\* as an  
assistant and each image transcription program into the binder is displayed  
(refer to drawing 20 ).

[0191] A user can make a key the thumbnail of the image transcription program  
by which it was indicated in this way by the list in the binder format, and can  
specify an image transcription program to reproduce. And what is necessary is  
just to read out a program to reproduce like "Channel G."

[0192] In the speech recognition section 13, speech recognition of the voice  
input from a user is carried out, and it is interpreted as a command. And while  
highlighting of the thumbnail of the image transcription program used as a target  
is carried out, it zooms in. (Refer to drawing 21 )

[0193] A user can give a key a thumbnail indication which zoomed in and can  
check that it is an image transcription program to reproduce. and playback  
initiation can be stopped in natural language formats, such as "be fastidious" and

"curve", and it can direct [ it can be alike and ]. In the speech recognition section 13, speech recognition of the voice input from a user is carried out, and it is interpreted as a command. And the voice as which means tofu lees mean Acknowledgements, such as "comprehension", is uttered (refer to drawing 22 ).

[0194] And while the thumbnail display of the image transcription program used as a target is expanded further and coming to occupy the whole monitor display soon, playback of the image transcription program concerned is started.

[0195] In addition, what is seen the whole date besides the selection format of an image transcription program which was mentioned above may be permitted. The scenario in this case is shown briefly [ below ].

[0196] a user -- "-- giving up -- 8 recorded yesterday -- showing -- "

-> Yesterday's video is turning to the circumference of a means ax (refer to drawing 25 ).

-> a user -- "-- Channel C -- showing -- "

-> It gives up and is "this ?."

-> User "it meets" (refer to drawing 26 )

-> The specified program display panel is expanded gradually and breadth and video are played by the whole screen (not shown).

[0197] (6) As shown in e-mail drawing 1 , it connects with the external network by the communication link interface 18 course, and the actuation support system 1

concerning this example can perform e-mail transmission and reception using the mail server on this network.

[0198] Moreover, the actuation support system 1 concerning this example can transmit by considering the personified assistant as agency and receive e-mail by the interaction of the voice base. Below, the case where e-mail is received supported by the voice base by the assistant is explained.

[0199] If e-mail arrives during program reception, the envelope icon which tells that e-mail is coming will fly down flutteringly on the image of a program (refer to drawing 27 ).

[0200] And if an envelope icon gets down to the lower limit of monitor display, the e-mail receiving icon which an envelope icon disappears and instead shows how many copies of mail was received will appear at the screen upper right. One copy of mail is received in the example shown in drawing 28 .

[0201] a user -- for example, -- "-- directions of the purport which opens a mail box can be given by giving up and asking an assistant in a natural language format like e-mail \*\*\*\*\*." That is, in the speech recognition section 13, speech recognition of the voice input from a user is carried out, and it is interpreted as a command. And means \*\* as an assistant and mail are opened. Furthermore, an assistant interprets the text described in e-mail, synthesizes voice, and may be made to read out.

[0202] In addition, in case e-mail is displayed on monitor display, kanji -> hiragana conversion is performed and you may enable it to read for children from the original text data. Moreover, those for old men in which the myopia force declined can be provided [ rather than ] with an easy operating environment in search of the alphabetic character of silk printing of a carbon button by voice control.

[0203] Moreover, the wallpaper of an e-mail viewing window may be customized for every user. For example, when a sending agency switches the wallpaper of e-mail according to a father, a mother, a child, a friend, etc., it can grasp at a glance who sent. Of course, you may make it switch a character font.

[0204] In the form of the flow chart shows the procedure for displaying the received mail on monitor display to drawing 30 . Hereafter, display processing of reception mail is explained according to this flow chart.

[0205] First, the text data contained in the body of reception mail are acquired (step S11).

[0206] Subsequently, the kanji contained in text data is changed into a hiragana (step S12). You may make it judge how much there is not necessarily no need of changing all kanjis into a hiragana, for example, it should be returned to a hiragana based on the user profile of age and others of a user here.

[0207] Subsequently, the text data by which hiragana conversion was carried out

are developed to bit map information using a font database (step S13). For example, two or more kinds of font databases are accumulated in the store 12. Moreover, you may make it choose the font database which should be used with reference to a user profile etc.

[0208] And the texture which serves as the so-called wallpaper in the text by which bit map expansion was carried out is superimposed, and the e-mail display which should be projected on monitor display is compounded (step S14). For example, two or more kinds of texture databases are accumulated in the store 12. Moreover, you may make it choose the texture which should be used with reference to a user profile etc.

[0209] (7) The message switching system using a computer like Message (notice plate) (Bulletin Board System) BBS is already established in information processing and the information communication link industry.

[0210] By this kind of conventional message and notice plate system, it is put on the condition that that message can be read from other users because each user writes a message in the notice plate which a specific server offers with the character input base.

[0211] On the other hand, the message and notice plate which the actuation support system 1 concerning this example offers can open a message to the input list of a message by the interaction of the voice base by considering the



personified assistant as agency. Moreover, when a user inputs a message, an assistant can read out a message only to a specific user by directing to an assistant that it is the message of addressing to a user of specification.

[0212] An example of the scenario of a message and a notice plate realized by the actuation support system 1 concerning this example is shown below (refer to drawing 31 ).

[0213] (Example)

Mother (user 1): "Do you give up and does a shopping line come just for a moment? Since it returns at 6:00"

(It kicks in mother appearance)

-- -- -- -- -- TAKUROU(user 2): -- "mama -- now -" (TAKUROU (child) goes home)

TAKUROU : "that ?" (he notices a mother's absence).

TAKUROU : "? which gives up and is not [ \*\*\*\*\* ]"

if it gives up and is : "mother (assistant) -- shopping -- \*\*\*\*\* . It will come back soon. "

TAKUROU: -- "\*\*\*\*\* -- is that right? Thank you"

[0214] (8) By the interaction of the voice base, the actuation support system 1 concerning service this example can form communication with a user by considering the personified assistant as agency.

[0215] The input from a user does not necessarily need to be a command to devices, such as a videocassette recorder by which external connection was made at the actuation support system 1 or this. For example, the interaction of a general conversational mode type may also be materialized. The interaction of this kind of user and an assistant is illustrated below.

[0216] (Example 1)

It is user: "tomorrow's weather is ?."

It gives up and is : "a candy (sad face)."

[0217] (Example 2)

User : "? which gives up and is carrying out central high-speed delay"

It gives up and is : "it is so so" (cool face).

[0218] (Example 3)

user: -- "-- giving up -- now -- when -- ?"

It gives up and is : "-- (nothing is said but a wrist watch is shown)."

[0219] (Example 4)

user: -- "-- giving up -- the San Francisco now -- when -- ?"

It gives up and is : "-- (nothing is said but a wrist watch is shown by the hand on which the glove of the Star-Spangled Banner was put.)."

[0220] (Example 5)

user: -- "-- giving up -- "

It gives up and is : "what."

user: -- "-- 6:00 of tomorrow -- starting -- "

giving up -- : "it is ? with what"

User: "there is an important meeting"

It gives up and is : "\*\*\*\*\*."

-- -- -- (television will be \*\*\*\* in 6:00 in in the next morning and the morning)

giving up -- : "good morning. Since there is an important meeting, it is \*\*\*\*?."

[0221] (Example 6)

user: -- "-- giving up -- \*\*\*\*\* -- "

If it gives up and is : "pizza store, it is still open. It orders. Is it a way? "

[0222] (9) The interactive actuation support system 1 concerning remote control

this example can perform the data exchange between an installation telephone

or a cellular phone via a general public line by modem 19 course

(above-mentioned). Similarly, since the system 1 is connected to wide area

networks, such as the Internet, by the communication link interface 18 course,

the information terminal of remote places, such as a personal computer, can

perform the data exchange.

[0223] Therefore, the personified assistant whom the actuation support system 1

offers can perform the interaction of the voice base between the terminals of

remote places, such as a cellular phone and a personal computer, and can form

communication with a user. For example, the actuation command to external instruments, such as a videocassette recorder, is receivable by the cellular-phone course.

[0224] However, when the user input from a remote place is received without any restriction, it may infringe on indoor security and privacy. Then, the actuation support system 1 requires the input of the authentication information for checking a user's justification from the information terminal of remote places, such as a cellular phone and a personal computer. A medium like an ID card may be used for the input of authentication information. Moreover, in an interior-of-a-room 1, i.e., actuation support system, side, voice and an entry of data are received via authentication equipment, a command interpretation is carried out and this is performed (refer to drawing 32 ).

[0225] In the actuation support system 1 concerning this example, an example of a scenario which receives the remote control from a user through the personified assistant is shown below (refer to drawing 33 - drawing 35 ).

[0226] (Example)

User : "? which gives up (telephoning a house from a cellular phone), and is" giving up -- : "it is"

user: -- "-- the NHK image transcription from 8:00 of today -- carrying out -- "

It gives up and he is NHK from : "8 o'clock."

User: "it meets so"

It gives up and is : "RYOKAI."

[0227] (10) The actuation support system 1 concerning acquisition this example of the text information from the program under broadcast contains the tuner 15 which performs, tuning, i.e., a channel selection, of the broadcast wave of a predetermined channel. The received broadcast wave is divided into a part for a part for image data division, and sound data division. The screen output of the image data is carried out through the image-processing section 16 at a monitor 25. Moreover, sound output of the sound data is carried out to a loudspeaker 21 through the speech synthesis section 14.

[0228] In the decoded image data, text, such as a title which accompanies this in addition to the volume on program book, may be included. For example, they are the score progress in sport relay broadcasts, such as professional baseball and soccer, a time stamp, etc.

[0229] For example, while broadcasting the sport in the hidden number group, there is a request of wanting to know the game progress.

[0230] In the actuation support system 1 concerning this example, the sports program as a hidden number group is decoded in a receiving list using the channel selection function in which the channel selection function of a tuner is opened while tuning in two or more preparations and a certain program. And the

text which carries out image measurement / recognition processing of the decoded image data, and corresponds to game progress is extracted. The personified assistant may be made to read out an extract result and you may make it notify it in the form of a title supermarket, a child window, etc. on a program current on display.

[0231] In the form of the flow chart shows the procedure for notifying a user of the text of a hidden number group to drawing 36 . Hereafter, notice processing of text is explained according to this flow chart.

[0232] First, score display area is extracted out of the image data of broadcast (step S21).

[0233] Subsequently, pattern matching is performed and score information is extracted from score display area (step S22).

[0234] The pattern match of the whole image data is carried out, and you may make it search for the corresponding text. However, since there is custom in which score display area is established in the almost same location as actually shown in drawing 37 , if this is made into a key, a score display can be discovered at high speed.

[0235] Subsequently, the extracted score information distinguishes whether it is changing from the last extract result (step S23).

[0236] If score information is changing, a user will be notified of this (step S24).

The gestalt of a notice may synthesize voice, for example and may be performed as an assistant's personified words. Or it may be made to perform a screen display using technique, such as a child window, 3D display, a texture, and 2D alpha blending.

[0237] On the other hand, if score information is not changing, it returns to step S21, and the same processing as \*\*\*\* is repeated and performed. In addition, score information may not be changing or you may make it notify a user of score information at intervals of predetermined time.

[0238] By the same technique as the above, a time stamp can be extracted out of a program image in addition to the score information on a sports program.

[0239] That is, when viewing and listening to the TV program as which time amount is displayed, or while carrying out the monitor of the hidden number group which performs a time amount display using the tuner which was equipped with two or more tuners and is vacant, time amount can be measured using image measurement / recognition technique.

[0240] For example, when remote connection of two or more external instruments is carried out through communication media, such as a home network, when local connection of two or more external instruments is made to the single actuation support system 1 or, it will become a complicated activity to double the time amount of all devices, although it is important. However,

malfunction may be invited if current time does not synchronize correctly between each device. For example, if the time of day by the side of a tuner is not exact in the case of program timed recording, an image transcription may go wrong.

[0241] While viewing and listening to the TV program in which the time amount display is performed in the actuation support system 1 concerning this example, or when having decoded the hidden number group by which it is indicated by time amount with the tuner as for which the tuner was equipped with two or more channel selection functions, and is vacant, image measurement / recognition processing is carried out, current time is measured and a television monitor 25 and the current time of an equipment when external connection of others was made are doubled automatically.

[0242] In the form of the flow chart shows the procedure for performing time amount doubling based on the time amount display in a program to drawing 38 . Hereafter, processing of time amount doubling is explained according to this flow chart.

[0243] First, time amount display area is extracted out of the image data of broadcast (step S31).

[0244] Subsequently, pattern matching is performed and a hour entry is extracted from time amount display area (step S32).



[0245] The pattern match of the whole image data is carried out, and you may make it search for the corresponding text. However, since there is custom in which time amount display area is established in the almost same location as actually shown in drawing 39 , if this is made into a key, a time amount display can be discovered at high speed.

[0246] Subsequently, the extracted time of day distinguishes whether it has changed to the last extract result (step S33).

[0247] If time of day has changed, this will be set up as time of day (step S34).

And the time of day of the external instrument to which others are connected further is set up (step S35).

[0248] On the other hand, if time of day has not changed, it returns to step S31, and the same processing as \*\*\*\* is repeated and performed.

[0249] It has explained in detail about this invention, referring to a specific example more than [addenda]. However, it is obvious that this contractor can accomplish correction and substitution of this example in the range which does not deviate from the summary of this invention.

[0250] Although this specification explained based on the operation gestalt which applied the interactive actuation support system applied to this invention to television actuation, the applicability of this invention is not limited to this. This invention can take effect similarly to a household-electric-appliances device and

information machines and equipment of the same kind equipped with the function which generates and displays the personified assistant, and voice input and recognition / composition function and the function to perform conversation with a user in a list with the voice base.

[0251] In short, with the gestalt of instantiation, this invention has been indicated and it should not be interpreted restrictively. In order to judge the summary of this invention, the column of the claim indicated at the beginning should be taken into consideration.

[0252]

[Effect of the Invention] As a full account was given above, according to this invention, the outstanding actuation support system and the outstanding actuation exchange approach of supporting the user command input to household-electric-appliances devices and information machines and equipment, such as a television set, can be offered.

[0253] Moreover, according to this invention, the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach of performing user command input to a device to a dialogue can be offered.

[0254] Moreover, according to this invention, the outstanding interactive actuation support system and the outstanding interactive actuation exchange

approach of inputting the user command to a device in a natural format through the assistant who personified can be offered.

[0255] Moreover, according to this invention, the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach the dialogue of the voice input base with the assistant who personified can perform command input can be offered.

[0256] Moreover, according to this invention, a user can offer the outstanding interactive actuation support system and the outstanding interactive actuation exchange approach of feeding back the advance situation of command actuation of having inputted by the dialogue of the voice input base with an assistant to a user.

---

## DESCRIPTION OF DRAWINGS

---

### [Brief Description of the Drawings]

[Drawing 1] It is drawing having shown typically the hardware configuration of the interactive actuation support system 1 with which operation of this invention is presented.

[Drawing 2] It is drawing having shown the command-processing system in the

actuation support system 1 concerning this example.

[Drawing 3] It is drawing having shown the character control system on the actuation support system 1 concerning this example.

[Drawing 4] It is drawing having shown the principle configuration for performing command processing in the voice base on the actuation support system 1 concerning this example.

[Drawing 5] It is drawing having shown the processing flow of character control roughly.

[Drawing 6] It is drawing having shown the example of the display screen just behind powering on of a television monitor 25.

[Drawing 7] It is drawing having shown signs that directions were given by the voice input of a natural language format to an assistant.

[Drawing 8] It is drawing having shown the example of a screen display when performing direct channel selection actuation through an assistant.

[Drawing 9] It is drawing having shown the example of a screen display when performing direct channel selection actuation through an assistant.

[Drawing 10] It is drawing having shown the example of a screen display when performing direct channel selection actuation through an assistant.

[Drawing 11] It is drawing having shown the example of a screen display when performing direct channel selection actuation through an assistant.

[Drawing 12] It is drawing having shown the procedure for realizing the user interface by direct command format.

[Drawing 13] It is drawing having shown the example of the multi-view screen which can command a whole view of the program under broadcast by each channel.

[Drawing 14] It is drawing having shown signs that each channel display panel rotated a circular ring top on a multi-view screen according to the directions from a user.

[Drawing 15] It is drawing having shown signs that it looked at the multi-view screen from the upper part virtually.

[Drawing 16] It is drawing in which Channel's C having changed into the temporary channel selection condition, and having shown signs that highlighting was carried out.

[Drawing 17] The channel selection of the channel C under temporary channel selection is decided, and the corresponding program display panel is drawing having shown signs that it expanded gradually.

[Drawing 18] It is drawing having shown signs that a program image transcription was directed with the voice base to an assistant.

[Drawing 19] It is drawing having shown the example of a screen display when performing playback actuation of an image transcription program through an

assistant.

[Drawing 20] It is drawing having shown the example of a screen display when performing playback actuation of an image transcription program through an assistant.

[Drawing 21] It is drawing having shown the example of a screen display when performing playback actuation of an image transcription program through an assistant.

[Drawing 22] It is drawing having shown the example of a screen display when performing playback actuation of an image transcription program through an assistant.

[Drawing 23] It is drawing having shown the example of a screen display in the case of setting up timed recording through an assistant.

[Drawing 24] It is drawing having shown the example of a screen display in the case of setting up timed recording through an assistant.

[Drawing 25] It is drawing having shown the example of a screen display in the case of specifying playback of an image transcription program the whole date through an assistant.

[Drawing 26] It is drawing having shown the example of a screen display in the case of specifying playback of an image transcription program the whole date through an assistant.

[Drawing 27] It is drawing having shown the interaction through the assistant when receiving e-mail.

[Drawing 28] It is drawing having shown the interaction through the assistant when receiving e-mail.

[Drawing 29] It is drawing having shown the interaction through the assistant when receiving e-mail.

[Drawing 30] It is the flow chart which showed the procedure for displaying the received mail on monitor display.

[Drawing 31] It is drawing having shown signs that agency of an assistant realized the message and the notice plate function.

[Drawing 32] It is drawing having shown structure for the actuation support system 1 concerning this example to receive the user input from the information terminal of a remote place.

[Drawing 33] In the actuation support system 1 concerning this example, it is drawing having shown signs that the remote control from a user was received through the personified assistant.

[Drawing 34] In the actuation support system 1 concerning this example, it is drawing having shown signs that the remote control from a user was received through the personified assistant.

[Drawing 35] In the actuation support system 1 concerning this example, it is

drawing having shown signs that the remote control from a user was received through the personified assistant.

[Drawing 36] It is the flow chart which showed the procedure for notifying a user of the text of a hidden number group.

[Drawing 37] It is drawing showing the location of the score display area in a program.

[Drawing 38] It is the flow chart which showed the procedure for performing time amount doubling based on the time amount display in a program.

[Drawing 39] It is drawing showing the location of the time amount display area in a program.

#### [Description of Notations]

1 -- Actuation support system

11 -- CC section

12 -- Storage

13 -- Speech recognition section

14 -- Speech synthesis section

15 -- Tuner

16 -- Image-processing section

17 -- Input/output interface

18 -- Communication link interface



19 -- Modem

21 -- Loudspeaker

22 -- Microphone

23 -- Videocassette recorder

24 -- Personal computer

25 -- Monitor

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2002-41276  
(P2002-41276A)

(43) 公開日 平成14年2月8日(2002.2.8)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テ-マ-コード(参考)
G 0 6 F 3/16	3 2 0	G 0 6 F 3/16	3 2 0 B 5 C 0 2 5
	3 3 0		3 3 0 E 5 C 0 5 6
3/00	6 0 1	3/00	6 0 1 5 C 0 6 3
G 1 0 L 15/00		H 0 4 N 5/00	A 5 D 0 1 5
15/22		5/44	Z 5 E 5 0 1

審査請求 未請求 請求項の数49 O L (全 41 頁) 最終頁に続く

(21) 出願番号 特願2000-222895(P2000-222895)

(22) 出願日 平成12年7月24日(2000.7.24)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 梨子田 辰志

東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

(72) 発明者 尾前 尚登

東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

(74) 代理人 100101801

弁理士 山田 英治 (外2名)

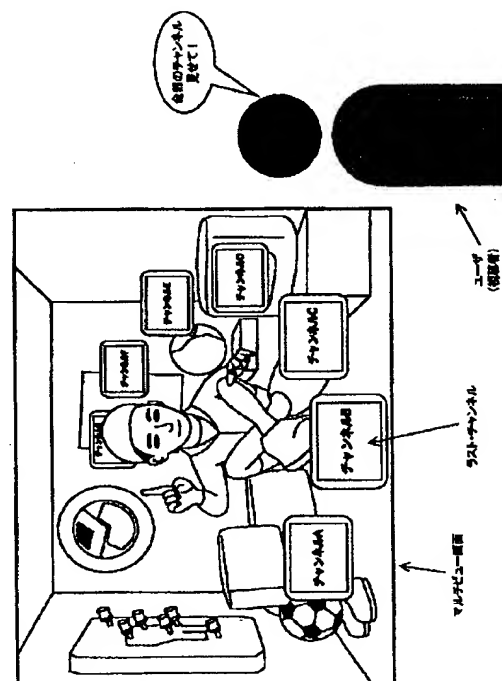
最終頁に続く

(54) 【発明の名称】 対話型操作支援システム及び対話型操作支援方法、並びに記憶媒体

## (57) 【要約】

【課題】 テレビ受像機などの家電機器や情報機器に対するユーザ・コマンド入力を支援する。

【解決手段】 操作支援システムによれば、音声合成やアニメーションによるリアクションを行なう擬人化されたアシスタントと呼ばれるキャラクタのアニメーションをユーザ・インターフェースとすることにより、ユーザに対して親しみを持たせると同時に複雑な命令への対応やサービスへの入り口を提供することができる。また、自然言語に近い感じの命令体系を備えているので、ユーザは、通常の会話と同じ感覚で機器の操作を容易に行なうことができる。



## 【特許請求の範囲】

【請求項1】機器に対するユーザ・コマンドの入力を支援する対話型操作支援システムであって、表示部と、音声入力部と、音声出力部と、演算制御部とを備え、前記演算制御部は、擬人化したアシスタントを生成して前記表示部に出現させるアシスタント制御手段と、該アシスタントが話す言葉を決定して音声合成して前記音声出力部を介して外部出力する出力音声制御手段と、前記音声入力部を介して入力されたユーザの声を音声認識する入力音声認識手段と、前記出力音声制御手段により決定されたアシスタントの言葉及び前記入力音声認識手段により認識されたユーザの言葉に従って両者間の対話を管理する対話管理手段と、前記対話管理手段が追跡した対話内容に基づいてユーザの意図又は入力コマンドを特定するコマンド解釈手段と、を具備することを特徴とする対話型操作支援システム。

【請求項2】前記アシスタント制御手段は、前記対話管理手段により管理される対話内容及び／又は前記コマンド解釈手段により特定されたユーザ入力コマンドに基づいて、アシスタントの妥当なアニメーションを決定することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項3】前記出力音声制御手段は、前記対話管理手段により管理される対話内容及び／又は前記コマンド解釈手段により特定されたユーザ入力コマンドに基づいて、該アシスタントが話す言葉を決定することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項4】前記出力音声制御手段は、前記コマンド解釈手段においてユーザの意図又は入力コマンドを特定できない場合には、ユーザの意図を導くような該アシスタントが話す言葉を決定することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項5】さらにテレビ受像機やビデオ・デッキなどの外部機器を接続する接続手段を備え、前記コマンド解釈手段は放送番組の選局やビデオ・デッキにおける録画／再生などの外部機器の機能コントロール用のコマンドも解釈することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項6】さらに外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段を備え、前記入力音声認識手段は、前記通信媒体経由で受信した音声データの認識処理も行うことを特徴とする請求項1に記載の対話型操作支援システム。

【請求項7】外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備え、前記出力音声制御手段は受信メールの内容に基づいてア

シスタントが話す言葉を決定することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項8】前記対話管理手段はあるユーザの入力音声を他のユーザへのメッセージとして管理し、前記出力音声制御手段は、該メッセージに基づいてアシスタントが話す言葉を決定することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項9】前記アシスタント制御手段は、擬人化されたアシスタントを情報資源へのリンクを含む各種のオブジェクトが散在する部屋に配置することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項10】前記アシスタント制御手段は、擬人化されたアシスタントを情報資源へのリンクを含む各種のオブジェクトが散在する部屋に配置し、該部屋に置かれた音楽コンテンツへのリンクを含む記録メディアに対してユーザが関心を抱いたことに応答して、前記コマンド解釈手段は該音楽コンテンツの再生としてコマンドの入力を解釈することを特徴とする請求項1に記載の対話型操作支援システム。

【請求項11】前記コマンド解釈手段がユーザからのコマンド入力を解釈したことに応答して、前記アシスタント制御手段はアシスタントを前記表示部に出現させることを特徴とする請求項1に記載の対話型操作支援システム。

【請求項12】さらにテレビ受像機を接続する接続手段を備え、

前記コマンド解釈手段がユーザからのチャンネル選局を入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、選局された放送番組の表示ウィンドウを携えたアシスタントを出現させることを特徴とする請求項1に記載の対話型操作支援システム。

【請求項13】さらにテレビ受像機を接続する接続手段を備え、

前記コマンド解釈手段がユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、切り替え可能な放送番組の各表示ウィンドウをアシスタントの周囲に略円環状に配置させることを特徴とする請求項1に記載の対話型操作支援システム。

【請求項14】さらに、ユーザによるチャンネル指示に追従して該略円環上で各表示ウィンドウを回転移動させ、チャンネル選局確定に応答して該当する放送番組表示ウィンドウをズームアップさせることを特徴とする請求項13に記載の対話型操作支援システム。

【請求項15】さらに放送番組コンテンツを蓄積・再生する2次記憶装置を接続する接続手段を備え、前記コマンド解釈手段がユーザからの録画番組の再生を入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、録画した放送番組を一覧表示したバインダを携えたアシスタントを出現させることを特徴

とする請求項 1 に記載の対話型操作支援システム。

【請求項 16】さらに、再生したい録画番組の選択確定に  
1 応答して該当する録画番組表示ウィンドウをズームアップさせることを特徴とする請求項 15 に記載の対話型操作支援システム。

【請求項 17】さらにテレビ受像機を接続する接続手段を備え、  
前記コマンド解釈手段がユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、切り替え可能な放送番組をマトリックス状に配置した放送番組一覧表を携えたアシスタントを出現させることを特徴とする請求項 1 に記載の対話型操作支援システム。

【請求項 18】さらに、チャンネル選局確定に応答して  
2 該当する放送番組表示ウィンドウをズームアップさせることを特徴とする請求項 17 に記載の対話型操作支援システム。

【請求項 19】データ放送の一部として配信される電子番組ガイド (EPG) を利用してマトリックス状の放送番組一覧表を生成することを特徴とする請求項 17 に記載の対話型操作支援システム。

【請求項 20】テレビ受像機を接続する接続手段と、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備え、  
前記アシスタント制御手段はメール受信に応答してメール受信表示を前記表示部上に出現させることを特徴とする請求項 1 に記載の対話型操作支援システム。

【請求項 21】さらに、前記表示部の画面上に表示するテキスト・データに関して漢字を仮名に変換する文字変換手段を備えることを特徴とする請求項 1 に記載の対話型操作支援システム。

【請求項 22】外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、  
該通信媒体経由で接続された情報端末の認証処理を行う認証手段と、をさらに備えることを特徴とする請求項 1 に記載の対話型操作支援システム。

【請求項 23】テレビ受像機を接続する接続手段と、受信させた放送番組コンテンツ中から文字情報を抽出する抽出手段と、をさらに備えることを特徴とする請求項 1 に記載の対話型操作支援システム。

【請求項 24】前記抽出手段によって抽出された文字情報を他の放送番組の表示番組上で表示することを特徴とする請求項 23 に記載の対話型操作支援システム。

【請求項 25】表示部と、音声入力部と、音声出力部とを備えた機器上で、当該機器又は外部接続された他の機器に対するユーザ・コマンドの入力を支援する対話型操作支援方法であって、  
擬人化したアシスタントを生成して前記表示部に出現させるアシスタント制御ステップと、

該アシスタントが話す言葉を決定して音声合成して前記音声出力部を介して外部出力する出力音声制御ステップと、

前記音声入力部を介して入力されたユーザの声を音声認識する入力音声認識ステップと、  
前記出力音声制御ステップにより決定されたアシスタントの言葉及び前記入力音声認識ステップにより認識されたユーザの言葉に従って両者間の対話を管理する対話管理ステップと、

10 前記対話管理ステップにより追跡された対話内容に基づいてユーザの意図又は入力コマンドを特定するコマンド解釈ステップと、を具備することを特徴とする対話型操作支援方法。

【請求項 26】前記アシスタント制御ステップでは、前記対話管理ステップにより管理される対話内容及び／又は前記コマンド解釈ステップにより特定されたユーザ入力コマンドに基づいて、アシスタントの妥当なアニメーションを決定することを特徴とする請求項 25 に記載の対話型操作支援方法。

20 【請求項 27】前記出力音声制御ステップでは、前記対話管理ステップにより管理される対話内容及び／又は前記コマンド解釈ステップにより特定されたユーザ入力コマンドに基づいて、該アシスタントが話す言葉を決定することを特徴とする請求項 25 に記載の対話型操作支援方法。

【請求項 28】前記出力音声制御ステップでは、前記コマンド解釈ステップにおいてユーザの意図又は入力コマンドを特定できない場合には、ユーザの意図を導くような該アシスタントが話す言葉を決定することを特徴とする請求項 25 に記載の対話型操作支援方法。

30 【請求項 29】前記機器はさらにテレビ受像機やビデオ・デッキなどの外部機器を接続する接続手段を備え、前記コマンド解釈ステップでは放送番組の選局やビデオ・デッキにおける録画／再生などの外部機器の機能コントロール用のコマンドも解釈することを特徴とする請求項 25 に記載の対話型操作支援方法。

【請求項 30】前記機器はさらに外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段を備え、

40 前記入力音声認識ステップでは、前記通信媒体経由で受信した音声データの認識処理も行うことを特徴とする請求項 25 に記載の対話型操作支援方法。

【請求項 31】前記機器は、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備え、

前記出力音声制御ステップでは受信メールの内容に基づいてアシスタントが話す言葉を決定することを特徴とする請求項 25 に記載の対話型操作支援方法。

50 【請求項 32】前記対話管理ステップではあるユーザの

入力音声进行他のユーザへのメッセージとして管理し、前記出力音声制御ステップでは、該メッセージに基づいてアシスタントが話す言葉を決定することを特徴とする請求項25に記載の対話型操作支援方法。

【請求項33】前記アシスタント制御ステップでは、擬人化されたアシスタントを情報資源へのリンクを含む各種のオブジェクトが散在する部屋に配置することを特徴とする請求項25に記載の対話型操作支援方法。

【請求項34】前記アシスタント制御ステップでは、擬人化されたアシスタントを情報資源へのリンクを含む各種のオブジェクトが散在する部屋に配置し、該部屋に置かれた音楽コンテンツへのリンクを含む記録メディアに対してユーザが関心を抱いたことに応答して、前記コマンド解釈ステップでは該音楽コンテンツの再生としてコマンドの入力を解釈することを特徴とする請求項25に記載の対話型操作支援方法。

【請求項35】前記コマンド解釈ステップにおいてユーザからのコマンド入力を解釈したことに応答して、前記アシスタント制御ステップはアシスタントを前記表示部に出現させることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項36】前記機器はさらにテレビ受像機を接続する接続手段を備え、前記コマンド解釈ステップにおいてユーザからのチャンネル選局を入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、選局された放送番組の表示ウィンドウを携えたアシスタントを出現させることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項37】前記機器はさらにテレビ受像機を接続する接続手段を備え、前記コマンド解釈ステップにおいてユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、切り替え可能な放送番組の各表示ウィンドウをアシスタントの周囲に略円環状に配置させることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項38】さらに、ユーザによるチャンネル指示に追従して該略円環上で各表示ウィンドウを回転移動させ、チャンネル選局確定に際して該当する放送番組表示ウィンドウをズームアップさせることを特徴とする請求項37に記載の対話型操作支援方法。

【請求項39】前記機器はさらに放送番組コンテンツを蓄積・再生する2次記憶装置を接続する接続手段を備え、

前記コマンド解釈ステップによりユーザからの録画番組の再生を入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、録画した放送番組を一覧表示したバインダを携えたアシスタントを出現させることを特徴とする請求項25に記載の対話型操作支援

方法。

【請求項40】さらに、再生したい録画番組の選択確定に際して該当する録画番組表示ウィンドウをズームアップさせることを特徴とする請求項39に記載の対話型操作支援方法。

【請求項41】前記機器はさらにテレビ受像機を接続する接続手段を備え、前記コマンド解釈ステップによりユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、切り替え可能な放送番組をマトリックス状に配置した放送番組一覧表を携えたアシスタントを出現させることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項42】さらに、チャンネル選局確定に際して該当する放送番組表示ウィンドウをズームアップさせることを特徴とする請求項41に記載の対話型操作支援方法。

【請求項43】データ放送の一部として配信される電子番組ガイド（EPG）を利用してマトリックス状の放送番組一覧表を生成することを特徴とする請求項41に記載の対話型操作支援方法。

【請求項44】前記機器は、テレビ受像機を接続する接続手段と、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備え、前記アシスタント制御ステップではメール受信に際してメール受信表示を前記表示部上に出現させることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項45】さらに、前記表示部の画面上に表示するテキスト・データに関して漢字を仮名に変換する文字変換ステップを備えることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項46】外部ネットワークや一般電話回線などの通信媒体と接続する通信ステップと、該通信媒体経由で接続された情報端末の認証処理を行う認証ステップと、をさらに備えることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項47】前記機器はテレビ受像機を接続する接続手段を備え、

受信させた放送番組コンテンツ中から文字情報を抽出する抽出ステップをさらに備えることを特徴とする請求項25に記載の対話型操作支援方法。

【請求項48】前記抽出ステップによって抽出された文字情報を他の放送番組の表示番組上で表示することを特徴とする請求項47に記載の対話型操作支援方法。

【請求項49】表示部と、音声入力部と、音声出力部とを備えた機器上で、当該機器又は外部接続された他の機器に対するユーザ・コマンドの入力を支援する対話型操作支援処理をコンピュータ・システム上で実行するように記述されたコンピュータ・ソフトウェアをコンピュー

タ可読形式で物理的に格納した記憶媒体であって、前記コンピュータ・ソフトウェアは、  
擬人化したアシスタントを生成して前記表示部に出現させるアシスタント制御ステップと、  
該アシスタントが話す言葉を決定して音声合成して前記音声出力部を介して外部出力する出力音声制御ステップと、  
前記音声入力部を介して入力されたユーザの声を音声認識する入力音声認識ステップと、  
前記出力音声制御ステップにより決定されたアシスタントの言葉及び前記入力音声認識ステップにより認識されたユーザの言葉に従って両者間の対話を管理する対話管理ステップと、  
前記対話管理ステップにより追跡された対話内容に基づいてユーザの意図又は入力コマンドを特定するコマンド解釈ステップと、を具備することを特徴とする記憶媒体。

#### 【発明の詳細な説明】

#### 【0001】

【発明の属する技術分野】本発明は、テレビ受像機などの家電機器や情報機器に対するユーザ・コマンド入力を支援する操作支援システム及び操作支援方法に係り、特に、各種の接続機器に対するユーザ・コマンド入力を対話式に行う対話型操作支援システム及び対話型操作支援方法に関する。

【0002】更に詳しくは、本発明は、擬人化したアシスタントを介して機器に対するユーザ・コマンドの入力を自然な形式で行うようにした対話型操作支援システム及び対話型操作支援方法に係り、特に、擬人化したアシスタントとの音声入力ベースの対話によりコマンド入力を行う対話型操作支援システム及び対話型操作支援方法に関する。

#### 【0003】

【従来の技術】従来、テレビ受像機などの各種家電機器、あるいは、パーソナル・コンピュータを始めとする情報機器に対するコマンド入力すなわちユーザ・インターフェースは、操作パネルや、キーボード、マウスなどのように、ユーザが手先で操作するタイプのものが主流であった。さらに、最近では、プロセッサの演算能力の向上や音声認識などの認知工学の進歩とも相俟って、音声入力ベースの対話型のユーザ・インターフェースも普及してきている。

【0004】前者の手先操作によるユーザ・インターフェースの場合、機器に対して直接・一義的なコマンドを入力することができるので操作が確実であるが、ユーザは機器の操作手法をある程度理解し習熟しなければならず、ユーザに課される負担が過大である。

【0005】例えば、テレビやその他のAV機器においては、メニューを十字キーなどでコントロールしていく「指先によるインターフェース」が中心である。このよ

うなユーザ入力様式では、ネットワーク家電では対応をとっていくのが煩雑となっていくことが予想される。

【0006】現在では、コマンダーを使用したユーザ・インターフェースが浸透しているが、地上波、衛星、インターネット、HAVI (Home Audio/Video Interoperability: デジタルAV機器の共通のコマンド体系) など、多チャンネル、多コントロールを行なおうとすると、スイッチの数が多くなり過ぎ、操作上困難になる。また、多機能スイッチ及びメニュー画面と組み合わせることにより、ある程度スイッチとの数を減らすことは可能となるが操作はむしろ煩雑になってしまう。

【0007】また、音声入力によるユーザ・インターフェースの場合、機器側において、入力音声の認識結果に基づいてユーザの要求を解析することによって、コマンドを特定することができるので、機器操作時のユーザに対する負担が軽減される。但し、ユーザは、話し相手が居ない場面でマイクに向かって話をしなければならず、人間の自然な行動とは言えず、対話入力に際して心理的な苦痛を伴うことさえある。

【0008】そこで、最近では、擬人化したアシスタントをディスプレイ・スクリーン上に出現させ、ユーザは、スクリーン上のアシスタントと対峙して会話を行う形式で機器に対するコマンド入力を実現するという対話型操作支援システムが出現してきている。

【0009】例えば、特開平11-65814号公報には、ユーザの発する音や音源方向を検出して、その検出結果に従って仮想生物の映像を制御することによって（例えば、仮想生物が音源に追従して常に声がする方向を見つめてくれるようにする）、ことによって、臨場感・現実感に富んだユーザ・インターフェースについて開示している。

【0010】また、特開平11-37766号公報には、車両内での運転者とのコミュニケーション機能を持つ擬人化されたエージェントを提供するエージェント装置について開示されている。該エージェント装置によれば、現在の車両・運転者の状況だけでなく、過去の履歴に基づく学習効果から擬人化されたエージェントが状況に合わせた行為をし、車両は運転者とのコミュニケーションを図ることができるようになっている。

【0011】最近では、計算機の演算能力の強化などによりハイレベルの対話処理が可能となり、画面上のアシスタントに対してもインテリジェンスを付与することが可能である。例えば、チャンネル選択や録画・再生の開始や停止などのように1単語で構成可能な単純な入力コマンドを操作するだけでなく、ユーザとの会話内容の文脈に追従して、複数の段階にまたがった複雑な操作を行うことが可能である。

【0012】しかしながら、このような操作の進行状況を、アシスタントが画面上で提示するようなシステムは未だ開発されておらず、ユーザは、ただ画面を見つめて

システムからの応答を待つしかない。応答時間を要する処理を命令したような場合には、機器の故障と誤解する場合さえある。

【0013】ユーザがアシスタントとの対話に基づいて機器を操作するためには、自然言語に近い感じの命令体系を操作上安易に入力できることが好ましい。

【0014】

【発明が解決しようとする課題】本発明の目的は、テレビ受像機などの家電機器や情報機器に対するユーザ・コマンド入力を支援することができる、優れた操作支援システム及び操作支援方法を提供することにある。

【0015】本発明の更なる目的は、機器に対するユーザ・コマンド入力を対話式に行うことができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することにある。

【0016】本発明の更なる目的は、擬人化したアシスタントを介して機器に対するユーザ・コマンドの入力を自然な形式で行うことができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することにある。

【0017】本発明の更なる目的は、擬人化したアシスタントとの音声入力ベースの対話によりコマンド入力を行うことができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することにある。

【0018】本発明の更なる目的は、ユーザがアシスタントとの音声入力ベースの対話により入力したコマンド操作の進行状況をユーザにフィードバックすることができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することにある。

【0019】

【課題を解決するための手段】本発明は、上記課題を参照してなされたものであり、その第1の側面は、機器に対するユーザ・コマンドの入力を支援する対話型操作支援システムであって、表示部と、音声入力部と、音声出力部と、演算制御部とを備え、前記演算制御部は、擬人化したアシスタントを生成して前記表示部に出現させるアシスタント制御手段と、該アシスタントが話す言葉を決定して音声合成して前記音声出力部を介して外部出力する出力音声制御手段と、前記音声入力部を介して入力されたユーザの声を音声認識する入力音声認識手段と、前記出力音声制御手段により決定されたアシスタントの言葉及び前記入力音声認識手段により認識されたユーザの言葉に従って両者間の対話を管理する対話管理手段と、前記対話管理手段が追跡した対話内容に基づいてユーザの意図又は入力コマンドを特定するコマンド解釈手段と、を具備することを特徴とする対話型操作支援システムである。

【0020】ここで、前記アシスタント制御手段は、前記対話管理手段により管理される対話内容及び／又は前記コマンド解釈手段により特定されたユーザ入力コマン

ドに基づいて、アシスタントの妥当なアニメーションを決定するようにしてもよい。

【0021】また、前記出力音声制御手段は、前記対話管理手段により管理される対話内容及び／又は前記コマンド解釈手段により特定されたユーザ入力コマンドに基づいて、該アシスタントが話す言葉を決定するようにしてもよい。

【0022】また、前記出力音声制御手段は、前記コマンド解釈手段においてユーザの意図又は入力コマンドを特定できない場合には、ユーザの意図を導くような該アシスタントが話す言葉を決定するようにしてもよい。

【0023】また、対話型操作支援システムは、さらにテレビ受像機やビデオ・デッキなどの外部機器を接続する接続手段を備えていてもよい。このような場合、前記コマンド解釈手段は放送番組の選局やビデオ・デッキにおける録画／再生などの外部機器の機能コントロール用のコマンドも解釈するようにしてもよい。

【0024】また、対話型操作支援システムは、さらに外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段を備えていてもよい。このような場合、前記入力音声認識手段は、前記通信媒体経由で受信した音声データの認識処理も行うようにしてもよい。

【0025】また、対話型操作支援システムは、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備えていてもよい。このような場合、前記出力音声制御手段は受信メールの内容に基づいてアシスタントが話す言葉を決定するようにしてもよい。

【0026】また、前記対話管理手段はあるユーザの入力音声他ユーザへのメッセージとして管理するようにしてもよい。このような場合、前記出力音声制御手段は、該メッセージに基づいてアシスタントが話す言葉を決定するようにしてもよい。

【0027】また、前記アシスタント制御手段は、擬人化されたアシスタントを情報資源へのリンクを含む各種のオブジェクトが散在する部屋（キャラクタ・ルーム）に配置するようにしてもよい。例えば、該部屋に置かれた音楽コンテンツへのリンクを含む記録メディアに対してユーザが関心を抱いたことに応答して、前記コマンド解釈手段は該音楽コンテンツの再生としてコマンドの入力を解釈するようにしてもよい。

【0028】また、前記コマンド解釈手段がユーザからのコマンド入力を解釈したことに応答して、前記アシスタント制御手段はアシスタントを前記表示部に出現させるようにしてもよい。

【0029】また、対話型操作支援システムは、さらにテレビ受像機を接続する接続手段を備えていてもよい。このような場合、前記コマンド解釈手段がユーザからのチャンネル選局を入力コマンドとして解釈したことに応

答して、前記アシスタント制御手段は、選局された放送番組の表示ウィンドウを携えたアシスタントを出現させるようにしてもよい。

【0030】あるいは、前記コマンド解釈手段がユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、切り替え可能な放送番組の各表示ウィンドウをアシスタントの周囲に略円環状に配置させるようにしてもよい。さらに、ユーザによるチャンネル指示に従って該略円環上で各表示ウィンドウを回転移動させ、チャンネル選局確定に応答して該当する放送番組表示ウィンドウをズームアップさせるようにしてもよい。

【0031】また、対話型操作支援システムは、さらに放送番組コンテンツを蓄積・再生する2次記憶装置を接続する接続手段を備えていてもよい。ここで言う2次記憶装置には、例えば、ビデオ・デッキや、ハード・ディスク、DVD-RAMドライブCD-R/Wドライブなど、大容量のメディア・コンテンツを記録可能なメディア記憶装置が含まれる。このような場合、前記コマンド解釈手段がユーザからの録画番組の再生を入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、録画した放送番組を一覧表示したバインダを携えたアシスタントを出現させるようにしてもよい。さらに、再生したい録画番組の選択確定に応答して該当する録画番組表示ウィンドウをズームアップさせるようにしてもよい。

【0032】また、対話型操作支援システムは、さらにテレビ受像機を接続する接続手段を備えていてもよい。このような場合、前記コマンド解釈手段がユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御手段は、切り替え可能な放送番組をマトリックス状に配置した放送番組一覧表を携えたアシスタントを出現させるようにしてもよい。さらに、チャンネル選局確定に応答して該当する放送番組表示ウィンドウをズームアップさせるようにしてもよい。また、データ放送の一部として配信される電子番組ガイド（EPG）を利用してマトリックス状の放送番組一覧表を生成するようにしてもよい。

【0033】また、対話型操作支援システムは、テレビ受像機を接続する接続手段と、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備えていてもよい。このような場合、前記アシスタント制御手段はメール受信に応答してメール受信表示を前記表示部上に表示させるようにしてもよい。

【0034】また、前記表示部の画面上に表示するテキスト・データに関して漢字を仮名に変換する文字変換手段をさらに備えていてもよい。

【0035】また、対話型操作支援システムは、外部ネットワークや一般電話回線などの通信媒体と接続するた

めの通信手段と、該通信媒体経由で接続された情報端末の認証処理を行う認証手段とをさらに備えていてもよい。

【0036】また、対話型操作支援システムは、テレビ受像機を接続する接続手段と、受信させた放送番組コンテンツ中から文字情報を抽出する抽出手段とをさらに備えていてもよい。このような場合、前記抽出手段によって抽出された文字情報を他の放送番組の表示番組上で表示するようにしてもよい。

10 【0037】また、本発明の第2の側面は、表示部と、音声入力部と、音声出力部とを備えた機器上で、当該機器又は外部接続された他の機器に対するユーザ・コマンドの入力を支援する対話型操作支援方法であって、擬人化したアシスタントを生成して前記表示部に表示させるアシスタント制御ステップと、該アシスタントが話す言葉を決定して音声合成して前記音声出力部を介して外部出力する出力音声制御ステップと、前記音声入力部を介して入力されたユーザの声を音声認識する入力音声認識ステップと、前記出力音声制御ステップにより決定されたアシスタントの言葉及び前記入力音声認識ステップにより認識されたユーザの言葉に従って両者間の対話を管理する対話管理ステップと、前記対話管理ステップにより追跡された対話内容に基づいてユーザの意図又は入力コマンドを特定するコマンド解釈ステップと、を具備することを特徴とする対話型操作支援方法である。

20 【0038】ここで、前記アシスタント制御ステップでは、前記対話管理ステップにより管理される対話内容及び／又は前記コマンド解釈ステップにより特定されたユーザ入力コマンドに基づいて、アシスタントの妥当なアニメーションを決定するようにしてもよい。

30 【0039】また、前記出力音声制御ステップでは、前記対話管理ステップにより管理される対話内容及び／又は前記コマンド解釈ステップにより特定されたユーザ入力コマンドに基づいて、該アシスタントが話す言葉を決定するようにしてもよい。

40 【0040】また、前記出力音声制御ステップでは、前記コマンド解釈ステップにおいてユーザの意図又は入力コマンドを特定できない場合には、ユーザの意図を導くような該アシスタントが話す言葉を決定するようにしてもよい。

【0041】また、前記機器がさらにテレビ受像機やビデオ・デッキなどの外部機器を接続する接続手段を備えるような場合には、前記コマンド解釈ステップでは放送番組の選局やビデオ・デッキにおける録画／再生などの外部機器の機能コントロール用のコマンドも解釈するようにしてもよい。

50 【0042】また、前記機器がさらに外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段を備えるような場合には、前記入力音声認識ステップでは、前記通信媒体経由で受信した音声データの認識処



理も行うようにしてもよい。

【0043】また、前記機器が、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備えるような場合には、前記出力音声制御ステップでは受信メールの内容に基づいてアシスタントが話す言葉を決定するようにしてもよい。

【0044】また、前記対話管理ステップではあるユーザの入力音声进行他のユーザへのメッセージとして管理し、前記出力音声制御ステップでは、該メッセージに基づいてアシスタントが話す言葉を決定するようにしてもよい。

【0045】また、前記アシスタント制御ステップでは、擬人化されたアシスタントを情報資源へのリンクを含む各種のオブジェクトが散在する部屋に配置するようにしてもよい。例えば、該部屋に置かれた音楽コンテンツへのリンクを含む記録メディアに対してユーザが関心を抱いたことに応答して、前記コマンド解釈ステップでは該音楽コンテンツの再生としてコマンドの入力を解釈するようにしてもよい。

【0046】また、前記コマンド解釈ステップにおいてユーザからのコマンド入力を解釈したことに応答して、前記アシスタント制御ステップはアシスタントを前記表示部に出現させるようにしてもよい。

【0047】また、前記機器がさらにテレビ受像機を接続する接続手段を備えるような場合には、前記コマンド解釈ステップにおいてユーザからのチャンネル選局を入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、選局された放送番組の表示ウィンドウを携えたアシスタントを出現させるようにしてもよい。

【0048】あるいは、前記コマンド解釈ステップにおいてユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、切り替え可能な放送番組の各表示ウィンドウをアシスタントの周囲に略円環状に配置させるようにしてもよい。さらに、ユーザによるチャンネル指示に追従して該略円環上で各表示ウィンドウを回転移動させ、チャンネル選局確定に応答して該当する放送番組表示ウィンドウをズームアップさせるようにしてもよい。

【0049】また、前記機器がさらに放送番組コンテンツを蓄積・再生するビデオ・デッキなどの2次記憶装置を接続する接続手段を備えるような場合には、前記コマンド解釈ステップによりユーザからの録画番組の再生を入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、録画した放送番組を一覧表示したバインダを携えたアシスタントを出現させるようにしてもよい。さらに、再生したい録画番組の選択確定に応答して該当する録画番組表示ウィンドウをズームアップさせるようにしてもよい。ここで言う2次記憶装置に

は、ビデオ・デッキ以外に、ハード・ディスクや、DVD-RAMドライブCD-R/Wドライブなど、大容量のメディア・コンテンツを記録可能なメディア記憶装置が含まれる。

【0050】また、前記機器はさらにテレビ受像機を接続する接続手段を備えるような場合には、前記コマンド解釈ステップによりユーザからのチャンネル切り替えを入力コマンドとして解釈したことに応答して、前記アシスタント制御ステップは、切り替え可能な放送番組をマトリックス状に配置した放送番組一覧表を携えたアシスタントを出現させるようにしてもよい。さらに、チャンネル選局確定に応答して該当する放送番組表示ウィンドウをズームアップさせるようにしてもよい。また、データ放送の一部として配信される電子番組ガイド(EPG)を利用してマトリックス状の放送番組一覧表を生成するようにしてもよい。

【0051】また、前記機器が、テレビ受像機を接続する接続手段と、外部ネットワークや一般電話回線などの通信媒体と接続するための通信手段と、該通信媒体経由で電子メールを交換するメール交換手段をさらに備えるような場合には、前記アシスタント制御ステップではメール受信に応答してメール受信表示を前記表示部に出現させるようにしてもよい。

【0052】また、対話型操作支援方法は、前記表示部の画面上に表示するテキスト・データに関して漢字を仮名に変換する文字変換ステップをさらに備えていてもよい。

【0053】また、対話型操作支援方法は、外部ネットワークや一般電話回線などの通信媒体と接続する通信ステップと、該通信媒体経由で接続された情報端末の認証処理を行う認証ステップとをさらに備えていてもよい。

【0054】また、前記機器がテレビ受像機を接続する接続手段を備える場合には、対話型操作支援方法は、受信させた放送番組コンテンツの中から文字情報を抽出する抽出ステップをさらに備えていてもよい。また、前記抽出ステップによって抽出された文字情報を他の放送番組の表示番組上で表示するようにしてもよい。

【0055】また、本発明の第3の側面は、表示部と、音声入力部と、音声出力部とを備えた機器上で、当該機器又は外部接続された他の機器に対するユーザ・コマンドの入力を支援する対話型操作支援処理をコンピュータ・システム上で実行するように記述されたコンピュータ・ソフトウェアをコンピュータ可読形式で物理的に格納した記憶媒体であって、前記コンピュータ・ソフトウェアは、擬人化したアシスタントを生成して前記表示部に出現させるアシスタント制御ステップと、該アシスタントが話す言葉を決定して音声合成して前記音声出力部を介して外部出力する出力音声制御ステップと、前記音声入力部を介して入力されたユーザの声を音声認識する入力音声認識ステップと、前記出力音声制御ステップによ

10

20

30

40

50

り決定されたアシスタントの言葉及び前記入力音声認識ステップにより認識されたユーザの言葉に従って両者間の対話を管理する対話管理ステップと、前記対話管理ステップにより追跡された対話内容に基づいてユーザの意図又は入力コマンドを特定するコマンド解釈ステップと、を具備することを特徴とする記憶媒体である。

【0056】本発明の第3の側面に係る記憶媒体は、例えば、様々なプログラム・コードを実行可能な汎用性のコンピュータ・システムに対して、コンピュータ・ソフトウェアをコンピュータ可読な形式で物理的に提供する媒体である。このような媒体は、例えば、CD (Compact Disc) やFD (Floppy Disc)、MO (Magneto-Optical disc) などの着脱自在で可搬性の記憶媒体である。あるいは、ネットワーク（ネットワークは無線、有線の区別を問わない）などの伝送媒体などを經由してコンピュータ・ソフトウェアを特定のコンピュータ・システムにコンピュータ可読形式で提供することも技術的に可能である。

【0057】このような記憶媒体は、コンピュータ・システム上で所定のコンピュータ・ソフトウェアの機能を実現するための、コンピュータ・ソフトウェアと記憶媒体との構造上又は機能上の協働的關係を定義したものである。換言すれば、本発明の第3の側面に係る記憶媒体を介して所定のコンピュータ・ソフトウェアをコンピュータ・システムにインストールすることによって、コンピュータ・システム上では協働的作用が発揮され、本発明の第1又は第2の各側面に係る対話型操作支援システム又は方法と同様の作用効果を得ることができる。

#### 【0058】

【作用】本発明に係る操作支援システム及び方法によれば、音声合成やアニメーションによるリアクションを行なう擬人化されたアシスタントと呼ばれるキャラクタのアニメーションをユーザ・インターフェースとすることにより、ユーザに対して親しみを持たせると同時に複雑な命令への対応やサービスへの入り口を提供することができる。また、自然言語に近い感じの命令体系を備えているので、ユーザは、通常の会話と同じ感覚で機器の操作を容易に行なうことができる。

【0059】本発明のさらに他の目的、特徴や利点は、後述する本発明の実施例や添付する図面に基づくより詳細な説明によって明らかになるであろう。

#### 【0060】

【発明の実施の形態】以下、図面を参照しながら本発明の実施例を詳解する。

【0061】図1には、本発明の実施に供される対話型操作支援システム1のハードウェア構成を模式的に示している。該システムは、例えば、セットトップボックス(STB)のようなテレビ受像機用の受信装置として構成され、テレビ受像機に接続されている。対話型操作支援システム1は、後述する「アシスタント」を媒介とし

てユーザとの対話を行ない、且つこの対話に基づきユーザの明示的又は潜在的な意図を解釈して、選局や録画・再生などのユーザ操作の支援サービスを提供することができる。以下、図1を参照しながら各部について説明する。

【0062】中央制御部11は、所定の制御プログラムに従って対話型操作支援システム1内の動作を統括的にコントロールする演算制御ユニットであり、例えば、ユーザのパートナーとしてのアシスタントの生成並びにアシスタントの行動パターンの発現と、音声及び画像の入出力をベースとしたユーザとアシスタントとの対話の管理などを行う。

【0063】中央制御部11において実行される制御プログラムには、例えば以下に示す機能が含まれる。すなわち、

(1) マイク22を介した音声入力及び音声認識により得られた入力コマンドに従って、システム1内の各部の動作制御。

(2) 入出力インターフェース17を介して接続された各種外部機器の制御

(3) チューナ15の制御

(4) アシスタントに関するキャラクタ・コントロール（音声認識で得られた入力コマンドに対応するアニメーションの生成）

(5) 音声合成（キャラクタが発する音声データの音声信号への変換。但し、キャラクタの口の動きなどのアニメーションと音声とが同期するようにする）

(6) 通信インターフェース18を介した外部ネットワークとの接続などの制御

(7) EPG (Electric Programming Guide) やその他のデータ放送用データの制御

(8) スピーカ21を介した音声出力のコントロール

(9) モニタ25を介した画面出力のコントロール。

(10) リモコン（図示しない）を介した入力コマンドによるコントロール

(11) 電子メール、EPG、広域ネットワーク上で使用されるテキスト・データの処理

(12) ユーザ・プロフィールに基づくテキスト・データの変換（例えば、お子様向けに「漢字」→「ひらがな」への変換）

(13) ビデオ信号を伴うデータを基にした画像計測（野球やサッカーなどのスポーツ番組上の得点経過や、画面表示からの時刻データの抽出）、画像認識に基づく各種サービス（得点経過の通知や時刻合わせなど）

(14) テキスト・データを基にフォント・データベースで選ばれたフォントを使用したビットマップ変換

(15) テクスチャ・データベースで選ばれたテクスチャと、フォントのビットマップとの合成

(16) システムの基本設定（画面の輝度、音量、各種入出力など）

【0064】チューナ15は、中央制御部11からの指示により、所定チャンネルの放送波のチューニングすなわち選局を行う。受信された放送波は、映像データ部分と音響データ部分に分離される。映像データは、画像処理部16を介してモニタ25に画面出力される。また、音響データは、音声合成部14を介してスピーカ21に音響出力される（あるいは、ライン出力であってもよい）。

【0065】記録装置12は、アシスタントのイメージ生成や行動パターンの生成のために必要なデータを保管している。保管データには、例えば以下のような情報が蓄積されている。すなわち、

- (1) アシスタントの3Dキャラクタ・イメージ情報やそのアニメーションを生成するために必要なデータ
- (2) アシスタントが活動するキャラクタ・ルームに関するレイアウトやその他の情報
- (3) アシスタントと対話するユーザのユーザ・プロフィール情報
- (4) ユーザとアシスタントとの過去の会話やその他のやり取りの履歴や履歴に基づく性格・感情や学習データ
- (5) アシスタントやキャラクタ・ルーム中に挿入すべき広告コンテンツ

【0066】また、記憶装置12は、アシスタントに関する情報以外にも、フォント・データベースやテキスト・データベースなどの各種のデータベース（図示しない）の保管も行なう。フォント・データベースは、例えば、EPGや掲示板、電子メールなどにおいて使用する各種フォントの管理を行なう。また、テキスト・データベースは、EPGや掲示板、電子メールなどにおいて使用する各種のテキスト（2Dのビットマップ・データなど）の管理を行なう。

【0067】システム1上では単一のアシスタントのみが出現するようにしてもよいが、ユーザ毎に異なるアシスタントを用意しておいてもよい。あらかじめ老若男女、性格などが相違する複数のキャラクタを用意しておき、初期ログイン時にユーザ・プロフィールに従って自動選択したりユーザ選択によりユーザ・プロフィールなどに関連付けて登録するようにしてもよい。また、アシスタントの学習・履歴データもユーザ毎に用意して、同じアシスタントであってもユーザ毎に別のリアクションをとるようにしてもよい。

【0068】音声認識部13は、マイク22などの音声入力装置を介した音声入力信号をユーザの音声をテキスト情報として認識して、さらに、言語データベース（図示しない）を利用しながら、テキスト形式に変換されたユーザ入力の解析処理を行う。より具体的には、形態素解析などによってテキストを単語単位に区分して、統語・意味談話解析によって統語情報・談話情報などの言語情報を取得して、ユーザ入力コマンドなどを理解して、中央制御部11に出力する。

【0069】入出力インターフェース17は、ビデオ・デッキ23やパーソナル・コンピュータ（PC）24などの外部機器を操作支援システム1に接続するための装置である。例えばi-link（IEEE1394）などのインターフェース仕様に従って1以上のAV機器や情報機器を外部接続することができる。

【0070】通信インターフェース18は、システム1を外部ネットワーク上の他のホスト計算機システムと相互接続するための装置である。外部ネットワークは、例えばインターネットのような広域ネットワークである。ネットワーク上には、HTML形式で記述されたWWW（World Wide Web）資源を提供するWWWサーバや、各ユーザ・アカウントに対してメール送受信サービスを提供するメール・サーバ、時々刻々更新される広告コンテンツを提供する広告サーバなどが散在する。本実施例では、ネットワーク上の少なくとも1つのサーバは、対話型操作支援を行うアシスタントに関するイメージやアニメーション、性格・行動モデルなどのキャラクタ・データを有料又は無料でディストリビュートするキャラクタ・サーバである。

【0071】この他、ネットワーク上には、公機関等のサービスを収集したSeason Database、天気予報や放送番組案内を週単位で提供するWeekly、ニュースや広告など即時性の高い情報を日単位で提供するDaily、株式市況や為替、交通情報などの時々刻々変動する情報を提供するTimelyなどの情報を提供サービスする情報提供サーバ、物流販売並びに決済（電子決済）をサービスするコマース・サーバ、インターネット・サービス・プロバイダなどが存在する。

【0072】例えばTCP/IP（Transmission Control Protocol/Internet Protocol）ネットワークであれば、各サーバにおいて提供される資源はURL（Uniform Resource Locator）形式で識別されるので、システム1は、HTTP（Hyper Text Transfer Protocol）などの所定プロトコルに従ってこれら情報資源をダウンロードすることができる。したがって、本実施例に係る操作支援システム1は、周期的又は任意のタイミングで、使用中のキャラクタ・データを再ダウンロードして、アシスタントのイメージや性格・機能などをアップデートすることができる。

【0073】ネットワーク経由で操作支援システム1が行ない得るトランザクションを以下に例示しておく。

(1) システム1内の各部を駆動する制御プログラムの更新

(2) アシスタントを構成するキャラクタのダウンロード

(3) フォント・データのダウンロード

(4) テキスト・データのダウンロード

(5) 番組録画したいときに録画する手段（例えばビデオ・テープなどのメディア）がないときの録画代行要求

(録画代行システムに関しては、本出願人に既に譲渡されている特開2000-162320号明細書を参照のこと)

(6) ユーザ・プロファイルの解析、ユーザ・カスタマイズ

(7) 公機関のサービス利用

(8) 天気予報、番組案内、ニュース、交通情報、広告などの取得

(9) 電子商取引

(10) ネットワーク経由でのキャラクタ・コントロール (音声、電子メール、制御用Webページなどを利用) 【0074】モデム19は、デジタル形式のコンピュータ・データをPSTN (Public Switched Telephone Network) などの一般公衆回線回線経由で伝送するための装置であり、アナログ信号への変調処理及びデジタル信号への復調処理を行う。

【0075】一般公衆回線には、例えば各家庭の設置電話が交換機接続されている。また、一般公衆回線は無線電話網とも相互接続されている。したがって、本実施例に係る対話型操作支援システム1は、設置電話や携帯電話との間でデータ交換を行なうことが可能である。また、中央制御部11において自動生成されるアシスタントは、携帯電話を介したユーザとの対話に基づいて入力コマンドの解釈を行なうことができる。

【0076】中央制御部11は、音声認識部13による音声認識結果に追従して、システム1内の動作をコントロールして、ユーザとシステム1間の対話を実現する。

【0077】例えば、ユーザ入力された音声のアシスタントとの会話である場合には、アシスタントに関する学習・履歴データや行動モデルに従って、アシスタントの挙動を決定して、音声や動画像によるアシスタントのリアクションを決定する。

【0078】アシスタントの音声は、音声合成部14により合成され、スピーカ41によって外部出力される。オンエア中の番組の音響が流れている場合には、これと重畳してアシスタントの音声を発するようにしてもよい。

【0079】また、アシスタントの動きは、3Dキャラクタ情報及びアニメーション情報を参照して、画像処理部14にて画像合成される。このとき、場合によってはキャラクタ・ルーム情報を参照して背景(シーン)を切り換えるようにしてもよい。また、アシスタントのイメージや、アシスタントが居るキャラクタ・ルームをオンエア中の1以上の番組映像と重畳表示するようにしてもよい(この点の詳細については後述に譲る)。

【0080】また、ユーザ入力された音声チャンネル切り換えである場合には、中央制御部11はチャンネル番号をチューナ15に転送して、選局動作に使用される。

【0081】また、ユーザ入力された音声、ビデオ・

デッキ23やパーソナル・コンピュータ24などの外部機器に対する指令(例えば録画や再生動作の開始・終了、早送り、巻き戻し、頭出し、ファイル転送)である場合には、中央制御部11は、入出力インターフェース17経由で該当する機器に対してコマンドを発行する。

【0082】また、ユーザ入力された音声、広域ネットワークへのアクセスを指示する場合には、通信インターフェース18経由でネットワーク上の指定されたホスト・システムに対するアクセス要求を転送する。例えば、ネットワークがインターネットなどのTCP/IPネットワークである場合には、アクセス要求はURL形式で記述される。このような場合、ユーザは、URLを読み上げてもよいし、あるいはURLと一意に関係付けられた言葉(例えばホームページのタイトル)を発してもよい。後者の場合、アシスタントを介して入力された音声は音声認識された後、URLに変換される。

【0083】なお、図示しないが、操作支援システム1は、従来のAV機器と同様、リモコンによるコマンド入力を受容してもよい。このような場合、リモコンから発信される無線(赤外線)データの受信部や受信データを解釈するデコーダを備え、デコードされたデータやコマンドは中央制御部11において演算処理される。

【0084】本実施例に係る操作支援システム1は、音声ベースでユーザからのコマンド入力を行なうが、その際、モニタ25の画面上に生成される「アシスタント」を仲介として対話的処理を行う点に特徴がある。ここで言うアシスタントは、アニメーション機能を備えた3Dキャラクタである。また、対話的処理は、ユーザとアシスタントとの(日常)会話を解釈してコマンドを抽出することや、アシスタントのリアクションを通してユーザ・フィードバックを与えることなどを含むものとする。

【0085】本実施例に係る操作支援システム1によれば、音声合成やアニメーションによるリアクションを行なう擬人化されたアシスタントをユーザ・インターフェースとすることにより、ユーザに対して親しみを持たせると同時に複雑な命令への対応やサービスへの入り口を提供することができる。また、操作支援システム1は、自然言語に近い感じの命令体系を備えているので、ユーザは、通常の会話と同じ感覚で機器の操作を容易に行なうことができる。

【0086】図2には、本実施例に係る操作支援システム1におけるコマンド処理体系を図解している。

【0087】マイクを介したユーザからの音声入力は、音声認識部13において、テキスト情報として認識して、さらに、言語データベース(図示しない)を利用しながら、テキスト形式に変換されたユーザ入力コマンドの解析処理が行われる。

【0088】本実施例では、例えば、ユーザ入力に呼応したアシスタントの挙動を生成するための「キャラクタ・コントロール」、操作支援システム1に接続されたビ

デオ・デッキその他の外部機器の動作指示を行なう「機器コントロール」、ユーザ入力に呼応したアシスタントの返答を生成する「音声合成」、ネットワークを介したメールの送受信を行う「メール」、複数の（不特定）ユーザ間でメッセージ交換を行なう「掲示板」などが、ユーザ入力可能なコマンドとして用意されている。

【0089】キャラクタ・コントロールは、ユーザからの音声ベースの入力コマンドに対して、アシスタントの挙動を伴って（言い換えれば、アシスタントの挙動をユーザ・フィードバックとして）、操作支援システム1又はこれに外部接続された機器のコントロールを行なうものである。

【0090】図3には、本実施例に係る操作支援システム1上におけるキャラクタ・コントロール体系を図解している。

【0091】同図に示すように、キャラクタ・コントロールを介して、「チャンネル選局」、「チャンネル切換」、「AV機器コントロール」、「メール読み」、「メール書き」、「掲示板読み」、「掲示板書き」、「Ambient」などの機能コマンドを実現することが

【0092】ここで、“Ambient”とは、待ち状態のときにアシスタントを構成するキャラクタが適当な動作をしたり、音声合成によりユーザ入力を催促するような機能を意味する。

【0093】また、「メール読み」や「掲示板読み」など、テキスト情報をキャラクタが読み上げる機能を実現する際には、自動的にリップシンク（発音の口の形）を行なうよう、音声とアニメーションの同期をとることが好ましい。

【0094】キャラクタ・コントロールを介した機能コマンドを実現する手順などについては後述に譲る。

【0095】キャラクタは、音声合成などの手段を介して現在の状態をユーザに通知するようにする。また、ユーザがネットワーク経由で、あるいは携帯電話を介して操作支援システム1を遠隔操作しているような場合には、電子メールなどの手段により状態通知を行なうようにしてもよい。

【0096】同一の操作支援システム1上では、アシスタントを構成する単一のキャラクタのみを用意しても、あるいは、ユーザ毎に個別にカスタマイズされたキャラクタを用意してもよい。各キャラクタのモデルやアニメーションは、例えば、ネットワーク、記録メディア、放送などの通信手段を介して、更新することができる。また、各キャラクタの衣装のテクスチャには、広告など他の情報コンテンツを動的にマッピングするようにしてもよい。

【0097】図4には、本実施例に係る操作支援システム1上で音声ベースでのコマンド処理を行なうための原理構成を図解している。

【0098】マイク22を介して入力された音声は、音声認識部13にて音声認識が施されてテキスト・ベースの情報に変換される。

【0099】中央制御部11は、テキスト情報を基に、ユーザとの対話処理を行ない、自然言語に近い形態でユーザから入力されたコマンドを理解する。

【0100】そして、中央制御部11は、コマンド解釈結果に従って、操作支援システム1に外部接続されたAV機器の動作をコントロールする。さらに、音声合成やキャラクタのアニメーションを生成したりして、指示コマンドの実行状況などをユーザ・フィードバックすることができる。

【0101】ここで、本実施例に係る操作支援システム1上で行なわれる音声入力ベースのコマンド解釈処理に関する幾つかの例を、以下に示しておく。

【0102】（例1）ユーザからの音声入力「ビデオ1をビデオ2にダビングして」

→ 入力された音声をテキストに変換し、さらに、「ビデオ」「1」「から」「ビデオ」「2」「に」「ダビング」「して」に字句分割する。

→ この命令形態を「ビデオ」「1」「2」という2つの機器と1つの命令「ダビング」に分類する。

→ これを字句解析して、「ビデオ」「1」「から」「ビデオ」「2」「に」「ダビング」という機器制御用のコマンドを生成する。

【0103】（例2）ユーザからの音声入力「ビデオをダビング」

→ 入力された音声を「ビデオ」「を」「ダビング」に字句分割する。

→ この命令形態には「ダビング」という命令が含まれているので、2つの機器機器があることが推定される。そこで、さらに何処から何処へという音声出力を行ない、これに対するユーザからの音声入力「ビデオ1からビデオ2へ」を受ける。

→ これを「ビデオ」「1」「から」「ビデオ」「2」「へ」と字句分割する。

→ 字句解釈に不足だったものを再入力して、ビデオ1からビデオ2へダビングという命令にする。

【0104】（例3）字句解釈には、同義語を登録して広範な言い回しに対処する。例えば、東京地区において「1channel」は、以下のすべての言い回しを網羅するものとする。

いち

いっちゃん

いっちゃんねる

総合

総合テレビ

えぬえっちけー総合

えぬえっちけー総合テレビ

50 えぬえっちけー

... ..

【0105】(例4)命令形態には、シングル、ダブル、トリプル…がある。シングル命令形態は、1つの単なる命令形態で、例えば「えぬえっちけー」や「よんちゃんねる」などである。また、ダブル命令形態は、1つの機器と1つの命令形態で構成される。例えば、テレビ録画に対する「テレビ」「録画」「して」(但し、「して」は省略可能である)や、録画したものの再生に対する「録画」「したものを」「見せて」などがダブル命令形態に相当する。また、トリプル命令形態は、2つの機器と1つの命令形態で構成される。例えば、ビデオ1からビデオ2へのダビングを指示する「ビデオ」「1」「から」「ビデオ」「2」「に」「ダビング」や、DVDからビデオ1へのダビングを指示する「DVD」「を」「ビデオ」「1」「に」「コピー」などがトリプル命令形態に相当する。

【0106】図5には、キャラクタ・コントロールの処理フローを概略的に示している。

【0107】例えば、テレビ・モニタ25の電源を投入することによって、アシスタントを構成するキャラクタは活動化する。アシスタントは、例えば、リビング・ルーム(又は自分の個室)内にたたずみ、「チャンネルの切換」、「動画録画」、「録画されたものを見る(再生)」、「メールを書く」、「メールを読む」、「Webを見る(WWW情報空間の探索)」、「伝言(掲示板への書き込み又は読み取り)」、「サービス」などのユーザ指示があるまで待機する。

【0108】ユーザは、アシスタントに対して、例えばマイク22を介した音声入力によって自然言語形式で指示を与えることができる。あるいは、外出先などから携帯電話を介してアシスタントに指示を与えることも許容されている。但し、リモート環境からは、チャンネル選択や録画番組の再生などの機能は意味をなさないで、「動画録画」、「メールを見る」、「サービス」など携帯電話に対して許可する機能を制限してもよい。

【0109】図6には、テレビ・モニタ25の電源投入直後の表示画面例を示している。

【0110】起動直後では、アシスタントは、“Ambient”、すなわち待ち状態であり、適当な動作をしたり、音声合成によりユーザ入力を催促するようなしぐさを行なう。

【0111】図示の例では、「よしお」と呼ばれるアシスタント・キャラクタがアシスタントの部屋(キャラクタ・ルーム)に出現している。各アシスタントは、デフォルト又はユーザ設定された性格を備えている。「よしお」には、例えば、以下に示すような条件が設定されている。

【0112】

【数1】よしおって…テレビの中に住んでいる元気な少年。日々、成長している。テレビをねぐらにしているか

ら、当然テレビのことなら何でも知っているし、テレビに繋がっているものなら、大抵のことはコントロールできる。好奇心旺盛だから、彼の部屋にはいつもいろんなものが置かれているし、最新の情報も、散らかり放題だ！ときどき訳のわからない行動もする。(まあ、こどもだからね)どうやら友達や仲間がいるらしいのだが、うまく逢えた人はまだ少ないとか…まあ、人のよいというか、かわいいインターフェースだと思ってくれ！！

【0113】よしおの部屋には、例えば、雑誌や玩具などの様々な物品(オブジェクト)が散在している。これらオブジェクトは、広告や物販(物理的なショッピングの他、データ・コンテンツや音楽の配信を含む)などの各種サービスへのリンク参照を有している。すなわち、室内には、アクセスできそうなものが散らばっており、ユーザが何かを言えば、どんな物事にアクセスできそうかを暗示している訳である。

【0114】アシスタントの部屋を用意することで、広告やその他の情報資源の利用が可能になる。また、アシスタントは、ユーザとの対話によりユーザのプロファイルを引き出すことで、より細かいサービスへの対応が可能になる。また、ユーザの言い回しや独特の癖などにも対応することができる。

【0115】例えば、ユーザは、よしおに対して興味のある物品のことを問い掛けることにより、音声認識及びコマンド解釈されて、物品オブジェクトがリンク参照する情報資源が呼び出される。室内に散在するオブジェクトやアシスタントの服装は、日ごとに変えてもよい。

【0116】また、操作支援システム1があらかじめ演奏可能な音楽コンテンツを蓄積しているような場合、図6に示すように、よしおの部屋にCDとCDプレーヤをおいてもよい。この場合、散在する各CD(記録メディア)は、該当する音楽コンテンツへのリンクを有している。そして、ユーザが「そのCD何?」とよしおに対して問いかけると、よしおが「〇△だよ。」と答えて(図7を参照のこと)、あたかも画面内のCDプレーヤで演奏するように、音楽コンテンツの再生を開始してもよい。

【0117】あるいは、よしおの部屋の床にヒットチャートのCDを置いておき、ユーザが「そのCD何?」とよしおに対して問いかけたことに対して、ユーザが当該ヒットチャート上の楽曲に興味を抱いていると判断して、CD購入(オンライン・ショッピング)などのリクエストを所定のサーバに発行したり、あるいは、所定サイトから当該音楽コンテンツのダウンロードを行なうようにしてもよい。

【0118】また、図6に示す例では、よしおの部屋の壁面上には、自動車のポスターが貼設されている。このように部屋に貼られたポスターは、広告媒体あるいは広告媒体へのリンク参照である。例えば、アシスタントによる発話形態で、購買誘導や広告情報の提示を行なうこ

とができる。

【0119】また、図6に示す例では、よしおの部屋には地球儀が設置されている。地球儀に対して、地球規模で展開された情報提供システムすなわちWWWというメタファを与えることができる。この場合、よしおに対して地球儀について話しかけることによって、WWWブラウザが起動して（図示しない）、WWW探索を行なうことができる。

【0120】また、図6に示す例では、よしおの部屋にはテレビ受像機が設置されている。これは、現実のテレビ・モニタ25についてのメタファを持っており、ラスト・チャンネル（すなわち最後に設定したチャンネル）の放映番組を表示するようにしてもよい（図示の例では、ラスト・チャンネルとしての「チャンネルB」が映し出されている）。

【0121】また、図6には示していないが、よしおの部屋にメール・ボックスを設置していてもよい。メール・ボックスは、電子メールの受信トレイについてのメタファを持っており、受信メールが蓄積されている場合には、メール・ボックス中に郵便物を表示するようにして、20 もよい。

【0122】なお、電源投入により直接このようなよしおの部屋の表示画面に向かうのではなく、ユーザによる発話やリモコン入力にตอบสนองしてよしおの部屋に辿りつくようにしてもよい。

【0123】本実施例に係る操作支援システム1によれば、よしおというアシスタントをユーザ・インターフェースとすることにより、ユーザに対して親しみを持たせると同時に複雑な命令への対応やサービスへの入り口を提供することができる訳である。

【0124】次いで、音声合成と3Dアニメーションによるリアクションをとるアシスタントを媒介とした、操作支援システム1による各種の操作支援処理について説明する。

#### 【0125】(1) テレビの電源ON/OFF

テレビの電源OFF状態（但し、操作支援システム1は稼動中）で、ユーザは、「よしおテレビ点けてよ」などのように自然言語形式で、アシスタントに問い掛ける。

【0126】このようなユーザ入力は音声認識部13において音声認識及びコマンド解釈され、テレビの電源が投入される。40

【0127】電源投入直後の初期画面は、例えば図6に示すようなよしおの部屋であってもよい。また、画面上でのよしおの登場の仕方も幾つかのバリエーションを用意しておき、ユーザ毎、あるいはユーザの状況や天候などに応じて適切なものを選択的に利用するようにしてもよい。アシスタントの登場の仕方としては、例えば以下のようなバリエーションが挙げられる。

【0128】・ よしおの部屋に「はい」、「じゃーん」とか言いながら、よしおが現れる。

・ 何も言わないで、よしおが部屋の中をうろうろしている。

・ 「あ」とか「えーとね」とか言いながら、うれしそうに待機（Ambient）状態に入る。

【0129】ここで、ユーザがさらに「テレビが見たい」旨の音声入力を行なうと、音声認識部13において音声認識及びコマンド解釈がなされる。そして、よしおは、「OK」あるいは「リョーカイ」などのような肯定応答を音声出力形式で行なうとともに、ラスト・チャンネル（すなわち前回の電源OFF時に選択されていたチャンネル）でテレビ映像が大きく映し出される。例えば、よしおの部屋に設置された仮想的なテレビ受像機にはラスト・チャンネルが映し出されており（前述）、この仮想的なテレビ・モニタ画面が徐々に拡大していき、やがて現実のモニタ画面全体を占めるように動作させてもよい。

【0130】他方、テレビの電源をOFFにするときには、ユーザは、「よしおテレビ消して」などのように自然言語形式でアシスタントに問い掛ければよい。

【0131】このようなユーザ入力は音声認識部13において音声認識及びコマンド解釈され、テレビの電源が遮断される。

【0132】また、音声入力を受け取ってからテレビの電源を遮断するまでの間、画面上にはよしおが現れて、少し悲しそうな顔をして背中を向けて画面奥行き方向に歩いて行き、画面から消失するという演出を行なうようにしてもよい。

#### 【0133】(2) チャンネルの選択

チャンネルの選択には、ユーザが特定のチャンネルを明示的に指示するダイレクトな選局と、現在放送中の番組の中から何か見たいというザッピング的な選局が挙げられる。

【0134】ダイレクトな選局：ダイレクトな選局においては、ユーザからの音声入力をベースに見たいチャンネルを特定することから、自然言語に近い感じの命令体系を操作上容易に入力できることが好ましい。

【0135】本実施例では、よしおというアシスタントをユーザ・インターフェースとすることにより、ユーザに対して親しみを持たせると同時に複雑な命令への対応やサービスへの入り口を提供することができる。例えば、最初の音声入力のみではユーザの意図を十分に解釈できないような場合には、分からないところを聞き返したりする擬人的なインタラクションにより、より自然なインターフェースを構築することができる。

【0136】ダイレクトな選局動作の例を以下に示しておく。

【0137】（例1）ユーザ「よしお、NHK見せて。1チャンネル」

→ よしお現れ「NHK？ 1チャン？」

→ ユーザ「そう（肯定応答）」



→ (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0138】(例2) ユーザ「よしお、教育見せて。3チャンネル」  
 → よしお現れ「教育? 3チャン?」  
 → ユーザ「そう(肯定応答)」  
 → (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0139】(例3) ユーザ「よしお、日テレ見せて。4チャンネル」  
 → よしお現れ「日テレ? 4チャン?」  
 → ユーザ「そう(肯定応答)」  
 → (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0140】(例4) ユーザ「よしお、TBS見せて。6チャンネル」  
 → よしお現れ「TBS? 6チャン?」  
 → ユーザ「そう(肯定応答)」  
 → (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0141】(例5) ユーザ「よしお、フジテレビ見せて。8チャンネル」  
 → よしお現れ「フジ? 8チャン?」  
 → ユーザ「そう(肯定応答)」  
 → (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0142】(例6) ユーザ「よしお、テレ朝見せて。10チャンネル」  
 → よしお現れ「テレ朝? 10チャン?」  
 → ユーザ「そう(肯定応答)」  
 → (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0143】(例7) ユーザ「よしお、テレビ東京見せて。12チャンネル」  
 → よしお現れ「テレビ東京? 12チャン?」  
 → ユーザ「そう(肯定応答)」  
 → (次の音声入力を待つ)  
 → (所定時間何もなかったことに応答して) よしお消える  
 【0144】図8~図11には、アシスタントを介してダイレクトな選局動作を行なうときの画面表示例を図解している。  
 【0145】まず、操作支援システム1のユーザは、テレビ番組の視聴者として放映中の野球中継番組を視聴し

ていたとする(図8を参照のこと)。このように放送受信中に、ユーザは、例えば、「よしお、8チャンネルにして」とか、「よしお、フジテレビにして」、あるいは「フジテレビ」のように、換えたいチャンネル名を明示的に言うことで、アシスタントを介してダイレクトな選局変更を行う。

【0146】このようなユーザからの音声入力は、音声認識部13において音声認識され、コマンドとして解釈される。そして、アシスタントとしてのよしおは、指示されたチャンネルで現在放送中の番組を表示したパネル(又はウィンドウ)を携えた状態で、モニタ画面上に出現する(図9を参照のこと)。この状態では、まだ切換先のチャンネルは仮選局の状態である。

【0147】そして、アシスタントが「これ?」などのセリフにより確認要求を行ない、これに対し、ユーザ側から「そう」などのような肯定応答が得られると、切換先のチャンネルは、仮選局の状態から選局確定の状態に移移する。

【0148】選局確定したことに応答して、アシスタントは、元のチャンネル番組(この例では野球中継)を表示するパネルを紙面右側に向かって押し出していくことにより、その後方に隠されていた切換先のチャンネル(この例では8チャンネル)で放映中の番組が徐々に現れてくる(図10及び図11を参照のこと)。

【0149】但し、元のチャンネルの番組を取り除く様式は、図10に示すような横方向にスライドさせるものに特に限定されない。例えば、アシスタントとしてのよしおが、元の番組のパネルを画面下方に向かって引き下げたり、あるいは、引き上げたりするようにしてもよい。

【0150】図8~図11は、単一のモニタ画面上に、元のチャンネル並びに切換先のチャンネルそれぞれで放送中の番組を同時に表示することで、ダイレクトな選局変更を視覚的に分かり易く表現している。2以上の放送番組の同時表示は、マルチプル・デコーディング機能により実現する。

【0151】なお、ダイレクトな選局変更のために登場するアシスタントの服などに、ロゴや広告情報を挿入することで、広告効果を期待することができる。勿論、アシスタントの服装を時節に応じて切り替えるようにしてもよい。

【0152】図12には、ダイレクト・コマンド形式によるユーザ・インターフェースを実現するための処理手順をフローチャートの形式で示している。該処理手順によれば、操作支援システム1は、ユーザからの音声入力のみでは分からない部分について、擬人化されたアシスタントを仲介として聞き返すというインタラクションにより、より自然なユーザ・インターフェースを実現することができる。以下、このフローチャートに従って説明する。

10

20

30

40

50



【0153】まず、マイク22を介したユーザからの音声入力を音声認識部13において音声認識して、入力キーワードを抽出する(ステップS1)。

【0154】次いで、カテゴリ毎に用意されたデータベースを検索することによって、該当するユーザ・コマンドを特定する(ステップS2)。カテゴリ別データベースは、例えば記憶装置12内に保管され、ネットワーク経由でその内容を更新することができる。

【0155】そして、現在までの音声入力のみでは不足で、ユーザ・コマンドを特定できない場合には、判断ブロックS3の分岐Noに進んで、アシスタントを媒介にして不足している情報を問い合わせる音声出力を行ない(ステップS5)、ステップS1に復帰して、ユーザからの次の音声入力を待つ。

【0156】他方、現在までの音声入力によりユーザ・コマンドを特定できた場合には、該当する機器(例えば、テレビ・モニタ25やビデオ・デッキ23など)にコマンドを発行する(ステップS4)。そして、ステップS1に復帰して、ユーザからの次の音声入力を待つ。

【0157】ザッピング的な選局：ザッピング的な選局を行いたい場合、ユーザは、「よしお全部見せて」とか「よしお何かやってる？」のような、敢えてチャンネルをチャンネルを特定しないような指示を、アシスタントに対して行なう。

【0158】このような音声指示に回答して、モニタ画面上にはアシスタントすなわちよしおが現れ、現在放送中の各番組を一望できるマルチビューを表示する。

【0159】図13には、各チャンネルで放送中の番組を一望できるマルチビュー画面の例を示している。同図に示す例では、アシスタントとしてのよしおの部屋の中で、各チャンネルで現在放送中の番組を表示するパネルが、よしおを取り囲むように円環状に配置されている。マルチビュー画面に切り替わった直後では、例えば、先頭にはラスト・チャンネル(すなわち最後に選局されたチャンネル)を配置するようにしてもよい。

【0160】このようなマルチビュー画面上では、ユーザは、アシスタントとしてのよしおに対して、自然言語形式で音声入力による指示を与えることができる。

【0161】例えば、「12チャンネル見せて」や「12チャンネル」、「テレビ東京見せて」、「テレビ東京」などのようなチャンネルを特定した指示を与えることによって、該当するチャンネルの番組表示パネルがモニタ画面上の最前面に来るように、各番組表示パネルを円環上で順次移動させる。また、「前」とか「次」などの指示により、各番組表示パネルを、円環上で1つずつ前方又は後方に移動させることができる(図14を参照のこと)。

【0162】図15には、マルチビュー画面を仮想的に上方から眺めた様子を示している。同図に示すように、よしおの部屋内には、番組表示パネルが移動する円環上

に沿って、各番組表示パネルが略等間隔で配置されている。また、円環の最前方には、現在仮選択されているチャンネルの番組表示パネルが配置される。

【0163】また、同図に示すように、円環の一部は番組表示パネルが配置されない間隙が設けられている。ユーザからのチャンネル移動指示に従って各番組表示パネル並びにこの間隙が円環に沿って順次移動するが、ユーザはこの間隙が移動した位置を目で追うことにより、チャンネルを移動させたことを視覚的にとらえることができる。

【0164】ある番組表示パネルが最前方に来たままの状態所で定時間(例えば数秒間)が経過すると、該当するチャンネルが仮選局状態となって、パネルがハイライト表示される。図16には、チャンネルCが仮選局状態となってハイライト表示されている様子を示している。

【0165】ユーザは、所望のチャンネルの番組表示パネルが画面の最前方に来たときに、「それ」、「これ」、「それ見せて」、「これ見せて」などのように強調して言うことによって、仮選局状態のチャンネルの選局を確定することができる。

【0166】選局が確定すると、該選択されたチャンネルのチューニングが行なわれるとともに、該チャンネルの番組表示パネルが徐々に拡大していき、やがてモニタ画面全体を占めるようになる。

【0167】図17には、仮選局中のチャンネルCの選局が確定して、該当する番組表示パネルが徐々に拡大していく様子を示している。

【0168】図13に示すようなマルチビュー画面では、単一のモニタ画面上に、各チャンネルで現在放送中の番組を同時に表示することで、ザッピング的な選局変更を視覚的に分かり易く表現している。2以上の放送番組の同時表示は、マルチプル・デコーディング機能により実現する。

【0169】なお、マルチビュー画面上に現れる番組表示パネルは、必ずしも現在放送中の番組である必要はなく、例えば、ビデオ・デッキやその他の記録装置上に蓄積された録画番組の中から取り出された再生番組であってもよい。

### 【0170】(3) 録画

本実施例に係る操作支援システム1は、外部機器として1台以上のビデオ・デッキ23を接続しており、受信した放送番組の録画先として指定することができる。また、ハード・ディスク・ドライブなどで構成される内部の記憶装置12を録画先として指定することができる。

【0171】また、本実施例に係る操作支援システム1は、擬人化されたアシスタントを仲介として、音声ベースのインタラクションにより、より自然なユーザ・インターフェースを提供することができる。したがって、リモコンなどの旧来の機器操作に頼ることなく、自然言語形式でアシスタントに問い掛けることによって、所望の

番組の録画指示を行なうことができる。

【0172】図18には、アシスタントに対して音声ベースで番組録画を指示する様子を示している。

【0173】同図(a)に示すように、ユーザは例えば野球中継の放送番組を視聴している。このように番組受信中に録画したくなった時点で、アシスタントに対して「録画して」とか「これ録画して」という音声ベースの指示を発する。

【0174】入力された音声は、音声認識部13において音声認識され、さらにコマンド解釈されて、システム1は入力コマンドが番組受信中の録画要求であることを特定する。

【0175】そして、外部接続されたビデオ・デッキ23又は内蔵ハード・ディスクなどに空き領域を探索する。録画先を確保した後、音声合成によりアシスタントを介して「リョーカイ」などの肯定応答を行なう。

【0176】また、図18(b)に示すように、モニタ画面上には録画中であることを示す録画アイコンと、録画時間を表すカウンタが出現する。

【0177】(4) 放送予定番組の録画

現在受信中の番組以外にも、放送予定番組の録画、すなわち予約録画を行なうこともできる。

【0178】本実施例に係る操作支援システム1は、擬人化されたアシスタントを仲介として音声ベースのインタラクションにより、より自然なユーザ・インターフェースを提供することができる。したがって、リモコンなどの旧来の機器操作に頼ることなく、自然言語形式でアシスタントに問い掛けることによって、予約録画の指示を行なうことができる。

【0179】図23及び図24には、アシスタントを介して予約録画を設定する場合の画面表示例を示している。

【0180】ユーザは、例えば、「よしおこれから何やるの?」とか「よしお次何やるの」、「よしおE P G (Electric Programming Guide) 見せてよ」、「よしお8時から何やるの」といったような、放送予定を引き出すようなせりふを、アシスタントを介して音声入力する。

【0181】このようなユーザからの音声入力、音声認識部13において音声認識され、コマンドとして解釈される。そして、アシスタントとしてのよしおは、マトリックス状に構成された放送予定の番組一覧を表示する(図23を参照のこと)。番組一覧は、例えばデータ放送用データとして配信されるE P Gを利用して生成することができる。

【0182】アシスタントは、図23に示すような番組一覧表をモニタ画面に提示した後、これを読み上げるようにしてもよい。

【0183】ユーザは、一覧表示された番組表の中から、録画したい番組を容易に見出すことができる。そし

て、例えば「8時からのチャンネルD録画しといて」などのような自然言語形式の指示を、アシスタントとしてのよしおを介して音声入力する。

【0184】このようなユーザからの音声入力は、音声認識部13において音声認識され、コマンドとして解釈される。そして、選択されたチャンネルDがハイライト表示される。また、アシスタントとしてのよしおは、チャンネルDの欄を指して、「これ?」などのように確認を促すせりふを音声出力する。

【0185】確認要求に対して、ユーザからの肯定的な応答があると、録画予約が設定される。この場合、録画指定された番組の欄をハイライト表示するだけでなく、録画予約アイコン(図示しない)を表示させるようにしてもよい。

【0186】(5) 録画番組の再生

本実施例に係る操作支援システム1は、擬人化されたアシスタントを仲介として音声ベースのインタラクションにより、より自然なユーザ・インターフェースを提供することができる。したがって、放送番組の録画だけでなく、録画番組の再生も、リモコンなどの旧来の機器操作に頼ることなく、自然言語形式でアシスタントに問い掛けることによって、所望の番組の録画指示を行なうことができる。

【0187】また、番組の録画先がハード・ディスク装置のようにランダム・アクセスが許容される記憶装置である場合には、任意の録画番組を取り出して再生を開始することができる。

【0188】図19～図22には、アシスタントを介して録画番組の再生動作を行うときの画面表示例を図解している。

【0189】まず、操作支援システム1のユーザは、テレビ番組の視聴者として放映中の歌謡番組を視聴していたとする(図19を参照のこと)。このように放送受信中に、ユーザは、「録ったやつ見せて」とか「録画したビデオ見せて」のように録画番組の再生を明示的に言うことで、アシスタントを介して再生指示を行うことができる。

【0190】このようなユーザからの音声入力は、音声認識部13において音声認識され、コマンドとして解釈される。そして、アシスタントとしてのよしおは、各録画番組の内容(番組のサムネイル)をバインダに入れたイメージを表示する(図20を参照のこと)。

【0191】ユーザは、このようにバインダ形式で一覧表示された録画番組のサムネイルを手がかりにして、再生したい録画番組を特定することができる。そして、再生したい番組を「チャンネルG」のように読み上げればよい。

【0192】ユーザからの音声入力は、音声認識部13において音声認識され、コマンドとして解釈される。そして、ターゲットとなる録画番組のサムネイルがハイラ

イト表示されるとともに、ズームアップされる。(図21を参照のこと)

【0193】ユーザは、ズームアップされたサムネイル表示を手がかりにして、再生したい録画番組であることを確認することができる。そして、「これ」とか「それ」などの自然言語形式で再生開始をよしおに指示することができる。ユーザからの音声入力、音声認識部13において音声認識され、コマンドとして解釈される。そして、よしおからは、「了解」などの確認応答を意味する音声が発せられる(図22を参照のこと)。

【0194】そして、ターゲットとなる録画番組のサムネイル表示がさらに拡大され、やがてモニタ画面全体を占めるようになるとともに、当該録画番組の再生が開始される。

【0195】なお、上述したような録画番組の選択様式以外にも、日にち毎で見ることを許容してもよい。この場合のシナリオを以下に簡単に示しておく。

【0196】ユーザ「よしお昨日録ったやつ見せて」  
→ よしおのまわりに昨日のビデオが回っている(図25を参照のこと)。

→ ユーザ「チャンネルCを見せて」

→ よしお「これ？」

→ ユーザ「そう」(図26を参照のこと)

→ 指定した番組表示パネルが徐々に拡大して画面全体に広がり、ビデオが再生される(図示しない)。

#### 【0197】(6) メール

図1に示すように、本実施例に係る操作支援システム1は、通信インターフェース18経由で外部のネットワークに接続されており、該ネットワーク上のメール・サーバを利用してメール送受信を行なうことができる。

【0198】また、本実施例に係る操作支援システム1は、擬人化されたアシスタントを仲介として音声ベースのインタラクションにより、メールの送受信を行なうことができる。以下では、アシスタントによる音声ベースでの支援によりメールを受信する場合について説明する。

【0199】番組受信中にメールが届くと、メールが来ていることを知らせる封筒アイコンが、番組の映像上にひらひらと舞い降りてくる(図27を参照のこと)。

【0200】そして、封筒アイコンがモニタ画面の下端まで降りてしまうと、封筒アイコンは消滅し、代わって、何通のメールが受信されたかを示すメール受信アイコンが画面右上に出現する。図28に示す例では、1通のメールが受信されている。

【0201】ユーザは、例えば、「よしお、メール見せて」などのように自然言語形式でアシスタントに問い掛けることで、メール・ボックスを開く旨の指示を与えることができる。すなわち、ユーザからの音声入力、音声認識部13において音声認識され、コマンドとして解釈される。そして、アシスタントとしてのよしおは、メ

ールをオープンする。さらに、アシスタントは、メール内に記述されたテキストを解釈して音声合成し、読み上げるようにしてもよい。

【0202】なお、メールをモニタ画面に表示する際、元のテキスト・データから、漢字→ひらがな変換を行ない、お子様向けに読めるようにしてもよい。また、近視力の衰えた老人向けに、音声コントロールによりボタンのシルク印刷の文字を探すよりも簡単な操作環境を提供することができる。

10 【0203】また、メール表示ウィンドウの壁紙をユーザ毎にカスタマイズしてもよい。例えば、発信元が父、母、子、友人などに応じてメールの壁紙を切り換えることによって、誰が発信したかを一目で把握することができる。勿論、文字フォントを切り換えるようにしてもよい。

【0204】図30には、受信したメールをモニタ画面に表示するための処理手順をフローチャートの形式で示している。以下、このフローチャートに従って受信メールの表示処理について説明する。

20 【0205】まず、受信メール本体に含まれるテキスト・データを取得する(ステップS11)。

【0206】次いで、テキスト・データ中に含まれる漢字をひらがなに変換する(ステップS12)。ここで、すべての漢字をひらがなに変換する必要は必ずしもなく、例えば、ユーザの年齢その他のユーザ・プロフィールに基づいて、どの程度ひらがなに戻すべきかを判断するようにしてもよい。

30 【0207】次いで、フォント・データベースを利用して、ひらがな変換されたテキスト・データをビットマップ情報に展開する(ステップS13)。例えば、記憶装置12には、複数種類のフォント・データベースが蓄積されている。また、ユーザ・プロフィールなどを参照して、使用すべきフォント・データベースを選択するようにしてもよい。

【0208】そして、ビットマップ展開されたテキストを、いわゆる壁紙となるテクスチャとを重畳して、モニタ画面に映し出すべきメール表示を合成する(ステップS14)。例えば、記憶装置12には、複数種類のテクスチャ・データベースが蓄積されている。また、ユーザ・プロフィールなどを参照して、使用すべきテクスチャを選択するようにしてもよい。

#### 【0209】(7) 伝言(掲示板)

BBS(Bulletin Board System)のようなコンピュータを利用したメッセージ交換システムは既に情報処理・情報通信業界において定着している。

【0210】従来の、この種の伝言・掲示板システムでは、特定のサーバが提供する掲示板に各ユーザがキャラクター入力ベースでメッセージを書き込むことで、他のユーザからもそのメッセージを読み得る状態に置かれるものである。

【0211】これに対し、本実施例に係る操作支援システム1が提供する伝言・掲示板は、擬人化されたアシスタントを仲介として音声ベースのインタラクションにより、メッセージの入力並びにメッセージの公開を行うことができる。また、ユーザがメッセージを入力するときに、特定のユーザ宛てのメッセージであることをアシスタントに対して指示することによって、アシスタントは特定のユーザに対してのみメッセージを読み上げるようにすることができる。

【0212】本実施例に係る操作支援システム1によって実現される伝言・掲示板のシナリオの一例を以下に示しておく(図31を参照のこと)。

【0213】(例)

母親(ユーザ1):「よしお、ちょっと買い物行ってくるね。6時に帰るから」  
(母親出かける)

… … … … …

タクロウ(ユーザ2):「ママ、ただいまー」(タクロウ(子供)が帰宅する)

タクロウ:「あれ?」(母親の不在に気付く)

タクロウ:「よしお、母さん知らない?」

よしお(アシスタント):「母さんなら買い物行ったよ。もうすぐ帰ってくるよ。」

タクロウ:「あっそうか。サンキュ」

【0214】(8) サービス

本実施例に係る操作支援システム1は擬人化されたアシスタントを仲介として音声ベースのインタラクションにより、ユーザとのコミュニケーションを成立させることができる。

【0215】ユーザからの入力、必ずしも操作支援システム1やこれに外部接続されたビデオ・デッキなどの機器に対するコマンドである必要はない。例えば、一般的な会話形式のインタラクションも成立し得る。この種のユーザとアシスタントとのインタラクションを以下に例示しておく。

【0216】(例1)

ユーザ:「明日の天気は?」

よしお:「あめ(悲しい顔)」

【0217】(例2)

ユーザ:「よしお、中央高速渋滞してる?」

よしお:「まあまあね(涼しい顔)」

【0218】(例3)

ユーザ:「よしお、いま何時?」

よしお:「…(何も言わず腕時計を見せる)」

【0219】(例4)

ユーザ:「よしお、サンフランシスコ今何時?」

よしお:「…(何も言わず、星条旗の手袋をはめた手で腕時計を見せる。)」

【0220】(例5)

ユーザ:「よしお」

よしお:「なあに」

ユーザ:「明日6時に起こして」

よしお:「何で?」

ユーザ:「大事な会議があるんだ」

よしお:「ふうん」

… … …

(翌朝、朝6時にテレビが点く)

よしお:「おはよう。大事な会議があるんでしょ?」

【0221】(例6)

ユーザ:「よしお、腹減ったよお」

よしお:「ピザ屋だったら、まだ開いているよ。オーダーしようか?」

【0222】(9) リモート・コントロール

本実施例に係る対話型操作支援システム1は、モデム19経由で一般公衆回線経由で設置電話や携帯電話との間でデータ交換を行なうことが可能である(前述)。同様に、システム1は通信インターフェース18経由でインターネットなどの広域ネットワークに接続されているので、パーソナル・コンピュータなどの遠隔地の情報端末ともデータ交換を行うことができる。

【0223】したがって、操作支援システム1が提供する擬人化されたアシスタントは、携帯電話やパーソナル・コンピュータなどの遠隔地の端末との間で音声ベースのインタラクションを行い、ユーザとのコミュニケーションを成立させることができる。例えば、携帯電話経由でビデオ・デッキなどの外部機器に対する操作コマンドを受信することができる。

【0224】但し、遠隔地からのユーザ入力を無制限に受け容れていると、室内のセキュリティやプライバシーが侵害されかねない。そこで、操作支援システム1は、携帯電話やパーソナル・コンピュータなどの遠隔地の情報端末に対しては、ユーザの正当性を確認するための認証情報の入力を要求する。認証情報の入力には、例えばIDカードのような媒体を利用してもよい。また、室内すなわち操作支援システム1側では、認証装置経由で音声やデータの入力を受け容れて、これをコマンド解釈して実行する(図32を参照のこと)。

【0225】本実施例に係る操作支援システム1において、擬人化されたアシスタントを介してユーザからのリモート・コントロールを受け容れるシナリオの一例を、以下に示しておく(図33~図35を参照のこと)。

【0226】(例)

ユーザ:「(携帯電話から自宅に電話して)よしおいる?」

よしお:「いるよ」

ユーザ:「今日8時からのNHK録画しといて」

よしお:「8時からのNHKだね」

ユーザ:「そうそう」

よしお:「リョーカイ!」

50 【0227】(10) 放送中の番組からのテキスト情報

## の取得

本実施例に係る操作支援システム1は、所定チャンネルの放送波のチューニングすなわち選局を行うチューナ15を含んでいる。受信された放送波は、映像データ部分と音響データ部分に分離される。映像データは、画像処理部16を介してモニタ25に画面出力される。また、音響データは、音声合成部14を介してスピーカ21に音響出力される。

【0228】デコードされた映像データ中には、番組本編以外に、これに付随する字幕などの文字情報を含むことがある。例えば、プロ野球やサッカーなどのスポーツ中継における得点経過や、時刻表示などである。

【0229】例えば、裏番組でスポーツを放送しているとき、その試合経過を知りたいという要望がある。

【0230】本実施例に係る操作支援システム1では、チューナの選局機能を複数備え、ある番組を選局中に開いている選局機能を用いて裏番組としてのスポーツ番組を受信並びにデコードする。そして、デコードされた映像データを画像計測・認識処理して試合経過に該当する文字情報を抽出する。抽出結果は、擬人化されたアシスタントに読み上げさせてもよいし、現在表示中の番組上で字幕スーパーや子ウィンドウなどの形式で通知するようにしてもよい。

【0231】図36には、裏番組の文字情報をユーザに通知するための処理手順をフローチャートの形式で示している。以下、このフローチャートに従って文字情報の通知処理について説明する。

【0232】まず、放送の映像データの中から、スコア表示エリアを抽出する（ステップS21）。

【0233】次いで、パターン・マッチングを行い、スコア表示エリアからスコア情報を抽出する（ステップS22）。

【0234】映像データ全体をパターンマッチして、該当する文字情報を探索するようにしてもよい。しかしながら、現実には図37に示すように、ほぼ同じ場所にスコア表示エリアが置かれる慣習があるので、これを手がかりにすれば、高速でスコア表示を発見することができる。

【0235】次いで、抽出したスコア情報が前回の抽出結果から変化しているか否かを判別する（ステップS23）。

【0236】スコア情報が変化していればユーザにこれを通知する（ステップS24）。通知の形態は、例えば音声合成して、擬人化されたアシスタントのセリフとして行ってもよい。あるいは、子ウィンドウ、3D表示、テクスチャ、2Dアルファブレンディングなど手法を用いて画面表示を行うようにしてもよい。

【0237】他方、スコア情報が変化していなければ、ステップS21に戻って上述と同様の処理を繰り返し実行する。なお、スコア情報が変化しなくても所定時

間間隔でスコア情報をユーザに通知するようにしてもよい。

【0238】上記と同様の手法により、スポーツ番組のスコア情報以外に、番組映像の中から時刻表示を抽出することができる。

【0239】すなわち、時間が表示されているテレビ番組を視聴しているときや、チューナを複数備えて空いているチューナを用いて時間表示を行う裏番組をモニタしているとき、画像計測・認識技術を用いて時間を計測することができる。

【0240】例えば、単一の操作支援システム1に対して複数の外部機器がローカル接続されているとき、あるいはホーム・ネットワークなどの通信媒体を介して複数の外部機器がリモート接続されているような場合、すべての機器の時間を合わせることは重要ではあるが煩雑な作業となってしまう。しかしながら、各機器間で現在時刻が正確に同期していないと、誤動作を招来する可能性がある。例えば、番組予約録画の際にチューナ側の時刻が正確でなければ録画に失敗することになりかねない。

【0241】本実施例に係る操作支援システム1では、時間表示が行われているテレビ番組を視聴しているとき、あるいは、チューナが複数の選局機能を備えて空いているチューナで時間表示されている裏番組をデコードしているとき、画像計測・認識処理して、現在時刻を計測してテレビ・モニタ25や、その他の外部接続された機器類の現在時刻を自動的に合わせるようになっている。

【0242】図38には、放送番組中の時間表示に基づいて時間合わせを行うための処理手順をフローチャートの形式で示している。以下、このフローチャートに従って時間合わせの処理について説明する。

【0243】まず、放送の映像データの中から、時間表示エリアを抽出する（ステップS31）。

【0244】次いで、パターン・マッチングを行って、時間表示エリアから時間情報を抽出する（ステップS32）。

【0245】映像データ全体をパターンマッチして、該当する文字情報を探索するようにしてもよい。しかしながら、現実には図39に示すように、ほぼ同じ場所に時間表示エリアが置かれる慣習があるので、これを手がかりにすれば、高速で時間表示を発見することができる。

【0246】次いで、抽出した時刻が前回の抽出結果と切り替わっているか否かを判別する（ステップS33）。

【0247】時刻が切り替わっていれば、これを時刻として設定する（ステップS34）。そして、さらにその他の接続されている外部機器の時刻を設定する（ステップS35）。

【0248】他方、時刻が切り替わっていなければ、ステップS31に戻って上述と同様の処理を繰り返し実行

10

20

30

40

50

する。

【0249】[追補]以上、特定の実施例を参照しながら、本発明について詳解してきた。しかしながら、本発明の要旨を逸脱しない範囲で当業者が該実施例の修正や代用を成し得ることは自明である。

【0250】本明細書では、テレビジョン操作に対して本発明に係る対話型操作支援システムを適用した実施形態に基づいて説明したが、本発明の適用範囲はこれに限定されるものではない。擬人化されたアシスタントを生成し表示する機能と、音声入力・認識・合成機能、並びに音声ベースでユーザとの会話を行う機能を備えた同種の家電機器・情報機器に対しても、同様に本発明は効果を奏することができる。

【0251】要するに、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本発明の要旨を判断するためには、冒頭に記載した特許請求の範囲の欄を参照すべきである。

#### 【0252】

【発明の効果】以上詳記したように、本発明によれば、テレビ受像機などの家電機器や情報機器に対するユーザ・コマンド入力を支援することができる、優れた操作支援システム及び操作支援方法を提供することができる。

【0253】また、本発明によれば、機器に対するユーザ・コマンド入力を対話式に行うことができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することができる。

【0254】また、本発明によれば、擬人化したアシスタントを介して機器に対するユーザ・コマンドの入力を自然な形式で行うことができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することができる。

【0255】また、本発明によれば、擬人化したアシスタントとの音声入力ベースの対話によりコマンド入力を行うことができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することができる。

【0256】また、本発明によれば、ユーザがアシスタントとの音声入力ベースの対話により入力したコマンド操作の進行状況をユーザにフィードバックすることができる、優れた対話型操作支援システム及び対話型操作支援方法を提供することができる。

#### 【図面の簡単な説明】

【図1】本発明の実施に供される対話型操作支援システム1のハードウェア構成を模式的に示した図である。

【図2】本実施例に係る操作支援システム1におけるコマンド処理体系を示した図である。

【図3】本実施例に係る操作支援システム1上におけるキャラクタ・コントロール体系を示した図である。

【図4】本実施例に係る操作支援システム1上で音声ベースでのコマンド処理を行なうための原理構成を示した図である。

【図5】キャラクタ・コントロールの処理フローを概略的に示した図である。

【図6】テレビ・モニタ25の電源投入直後の表示画面例を示した図である。

【図7】アシスタントに対して自然言語形式の音声入力により指示を与える様子を示した図である。

【図8】アシスタントを介してダイレクトな選局動作を行なうときの画面表示例を示した図である。

【図9】アシスタントを介してダイレクトな選局動作を行なうときの画面表示例を示した図である。

【図10】アシスタントを介してダイレクトな選局動作を行なうときの画面表示例を示した図である。

【図11】アシスタントを介してダイレクトな選局動作を行なうときの画面表示例を示した図である。

【図12】ダイレクト・コマンド形式によるユーザ・インターフェースを実現するための処理手順を示した図である。

【図13】各チャンネルで放送中の番組を一望できるマルチビュー画面の例を示した図である。

【図14】マルチビュー画面上で、ユーザからの指示に従って各チャンネル表示パネルが円環上を回転移動する様子を示した図である。

【図15】マルチビュー画面を仮想的に上方から眺めた様子を示した図である。

【図16】チャンネルCが仮選局状態となってハイライト表示されている様子を示した図である。

【図17】仮選局中のチャンネルCの選局が確定して、該当する番組表示パネルが徐々に拡大していく様子を示した図である。

【図18】アシスタントに対して音声ベースで番組録画を指示する様子を示した図である。

【図19】アシスタントを介して録画番組の再生動作を行うときの画面表示例を示した図である。

【図20】アシスタントを介して録画番組の再生動作を行うときの画面表示例を示した図である。

【図21】アシスタントを介して録画番組の再生動作を行うときの画面表示例を示した図である。

【図22】アシスタントを介して録画番組の再生動作を行うときの画面表示例を示した図である。

【図23】アシスタントを介して予約録画を設定する場合の画面表示例を示した図である。

【図24】アシスタントを介して予約録画を設定する場合の画面表示例を示した図である。

【図25】アシスタントを介して日にち毎で録画番組の再生を指定する場合の画面表示例を示した図である。

【図26】アシスタントを介して日にち毎で録画番組の再生を指定する場合の画面表示例を示した図である。

【図27】メールを受信したときのアシスタントを介したインタラクションを示した図である。

【図28】メールを受信したときのアシスタントを介し

たインタラクションを示した図である。

【図29】メールを受信したときのアシスタントを介したインタラクションを示した図である。

【図30】受信したメールをモニタ画面に表示するための処理手順を示したフローチャートである。

【図31】アシスタントの仲介により伝言・掲示板機能を実現した様子を示した図である。

【図32】本実施例に係る操作支援システム1が遠隔地の情報端末からのユーザ入力を受け容れるための仕組みを示した図である。

【図33】本実施例に係る操作支援システム1において、擬人化されたアシスタントを介してユーザからのリモート・コントロールを受け容れる様子を示した図である。

【図34】本実施例に係る操作支援システム1において、擬人化されたアシスタントを介してユーザからのリモート・コントロールを受け容れる様子を示した図である。

【図35】本実施例に係る操作支援システム1において、擬人化されたアシスタントを介してユーザからのリモート・コントロールを受け容れる様子を示した図である。

【図36】裏番組の文字情報をユーザに通知するための\*

\* 処理手順を示したフローチャートである。

【図37】放送番組中におけるスコア表示エリアの場所を表した図である。

【図38】放送番組中の時間表示に基づいて時間合わせを行うための処理手順を示したフローチャートである。

【図39】放送番組中における時間表示エリアの場所を表した図である。

#### 【符号の説明】

1…操作支援システム

10 11…中央制御部

12…記憶装置

13…音声認識部

14…音声合成部

15…チューナ

16…画像処理部

17…入出力インターフェース

18…通信インターフェース

19…モデム

21…スピーカ

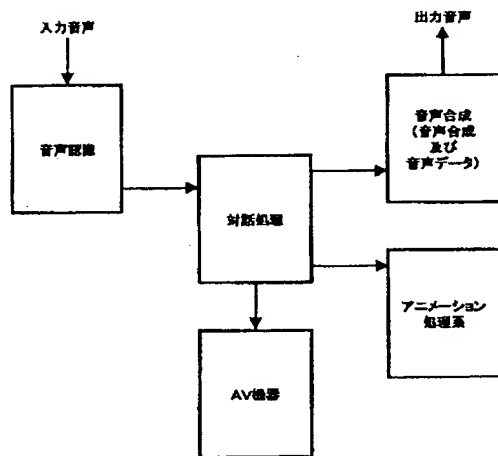
20 22…マイク

23…ビデオ・デッキ

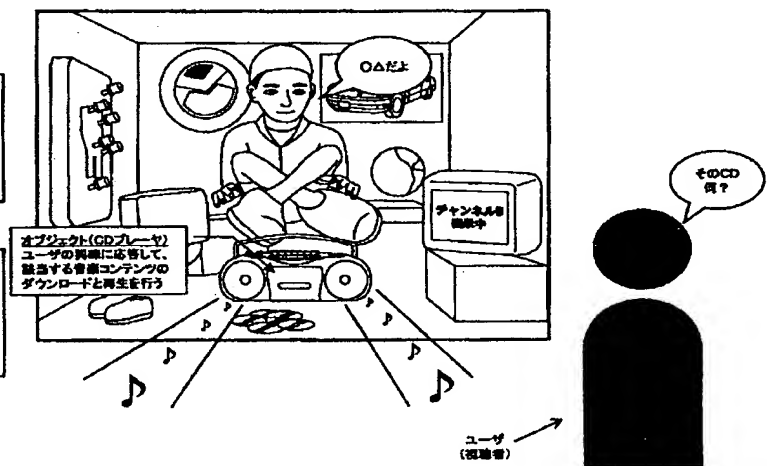
24…パーソナル・コンピュータ

25…モニタ

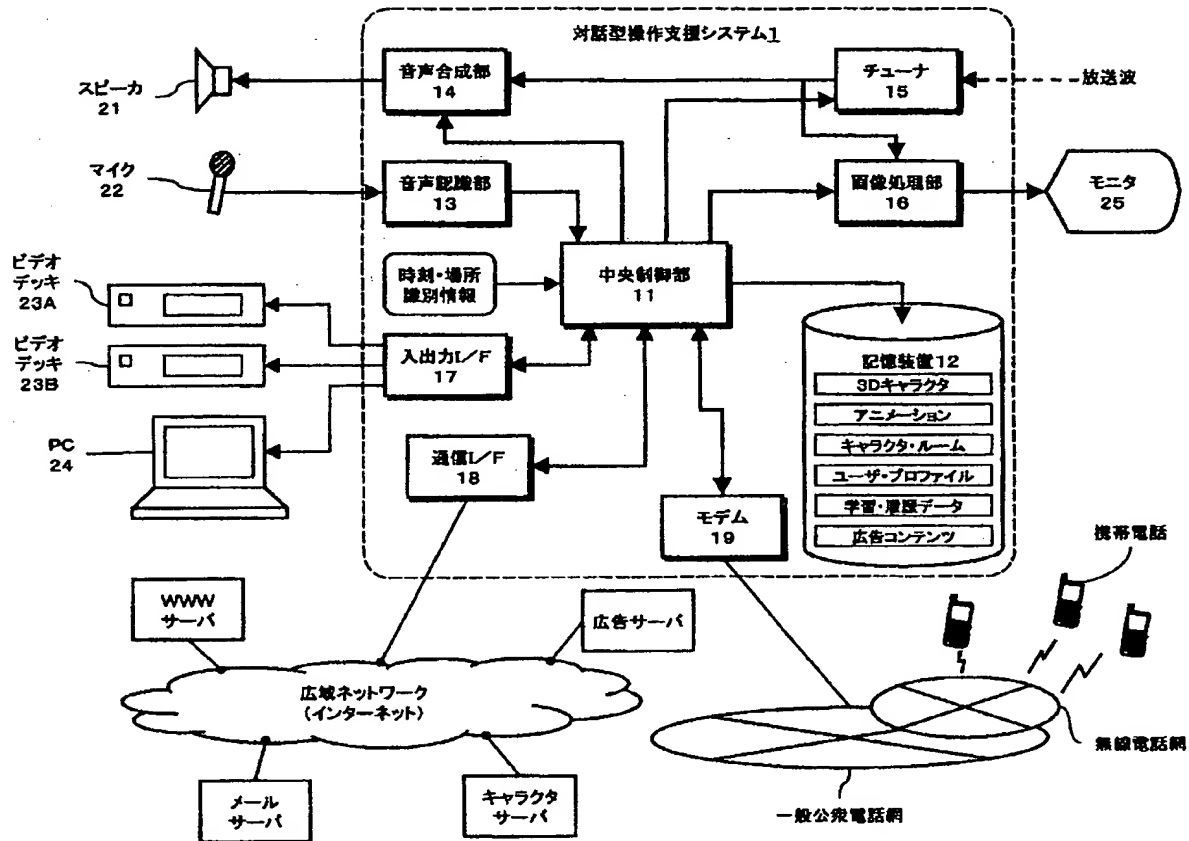
【図4】



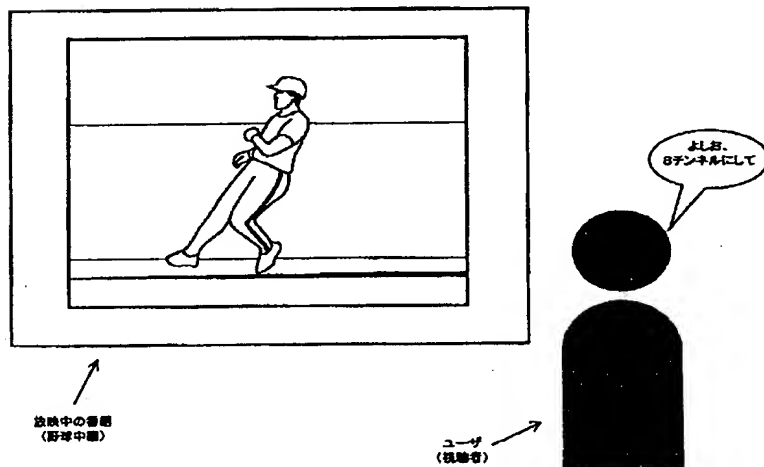
【図7】



【図1】

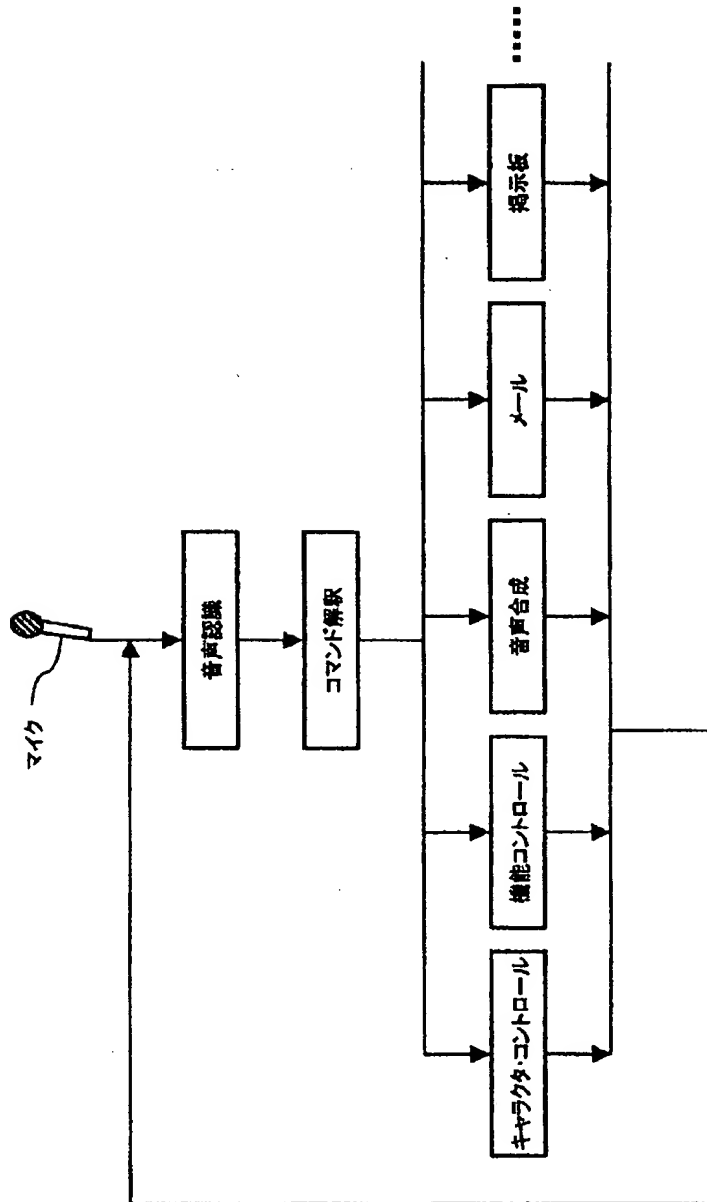


【図8】

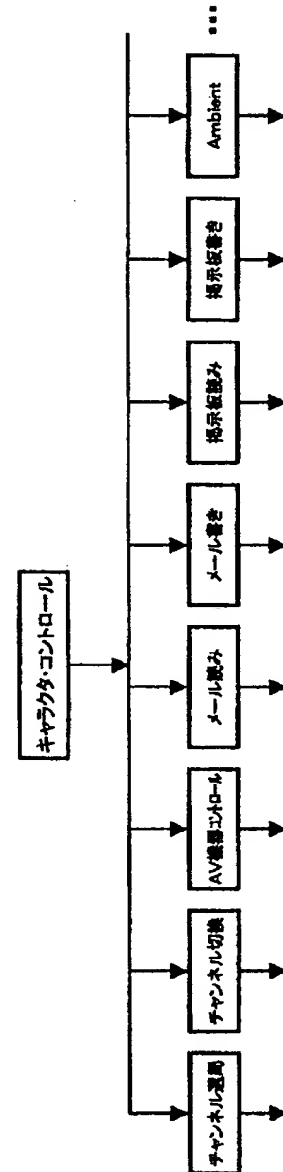




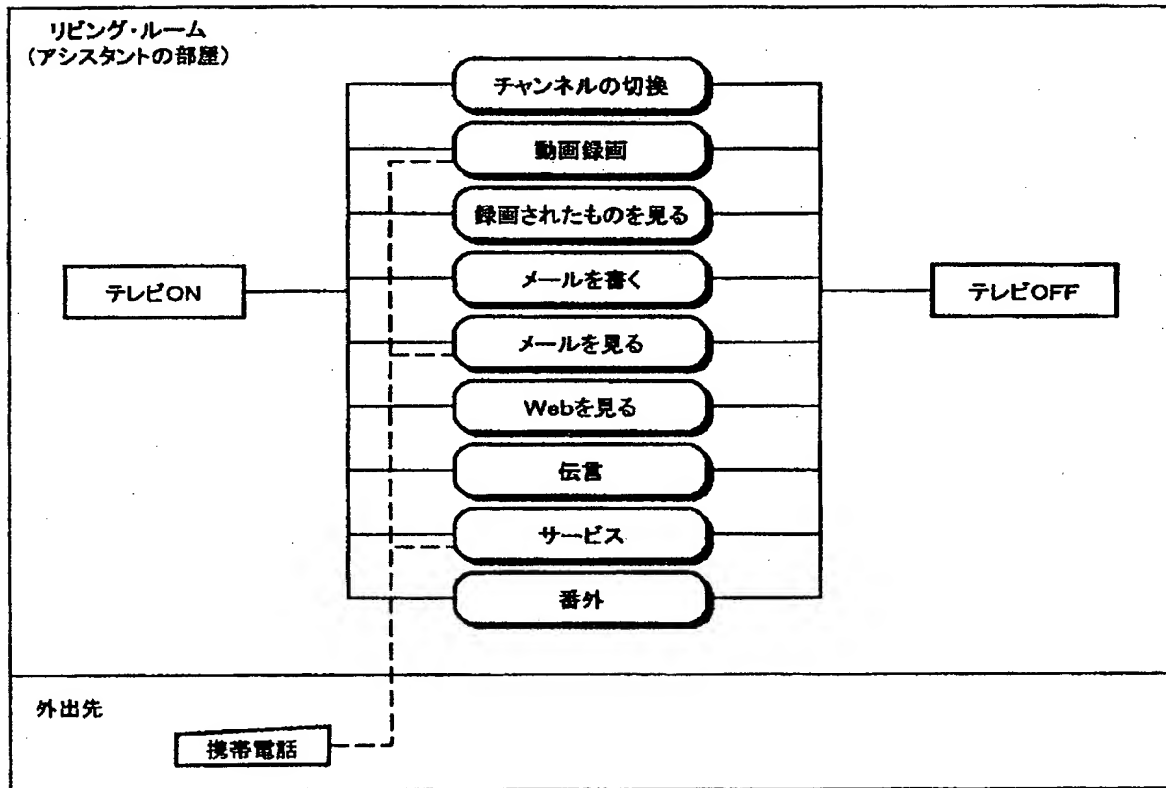
【図2】



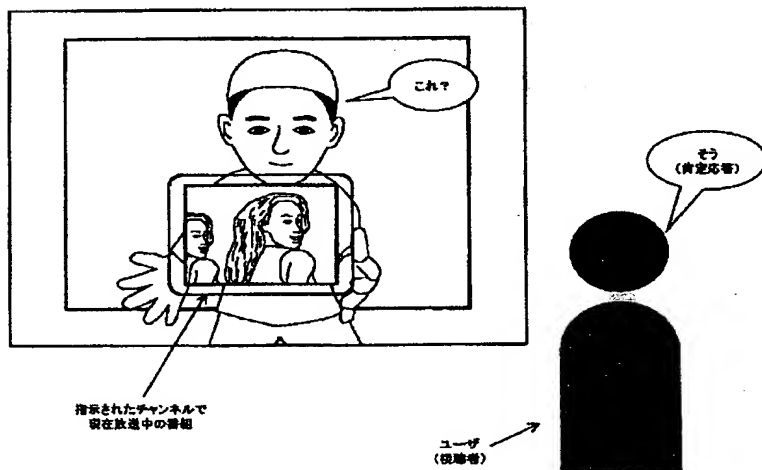
【図3】



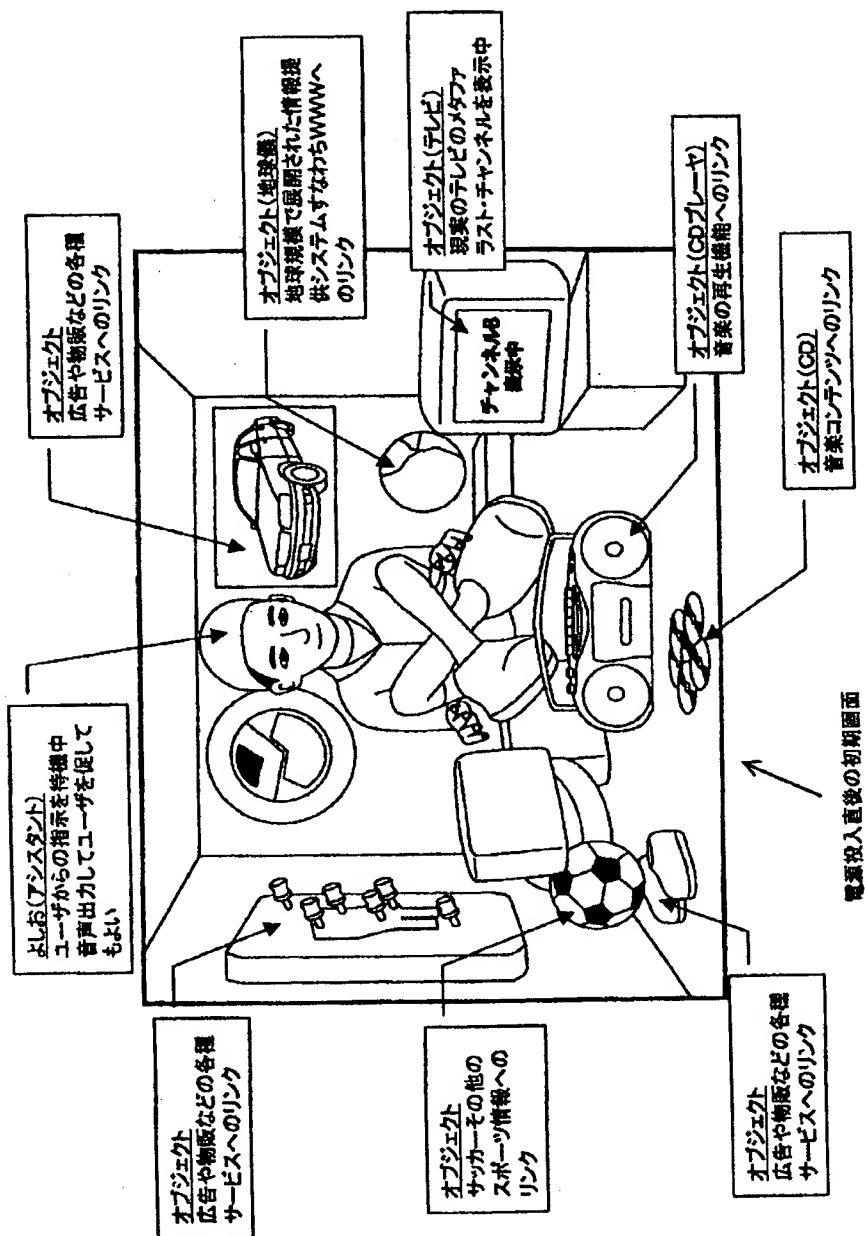
【図5】



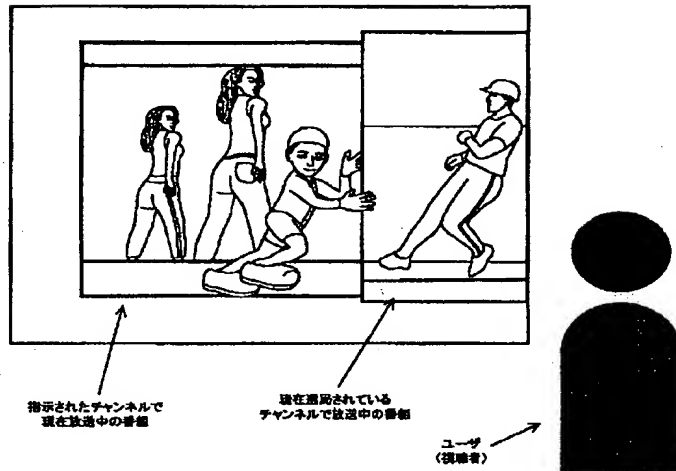
【図9】



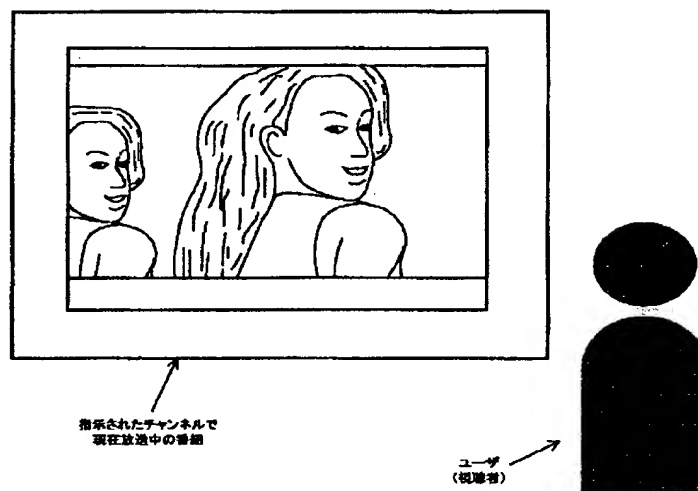
【図6】



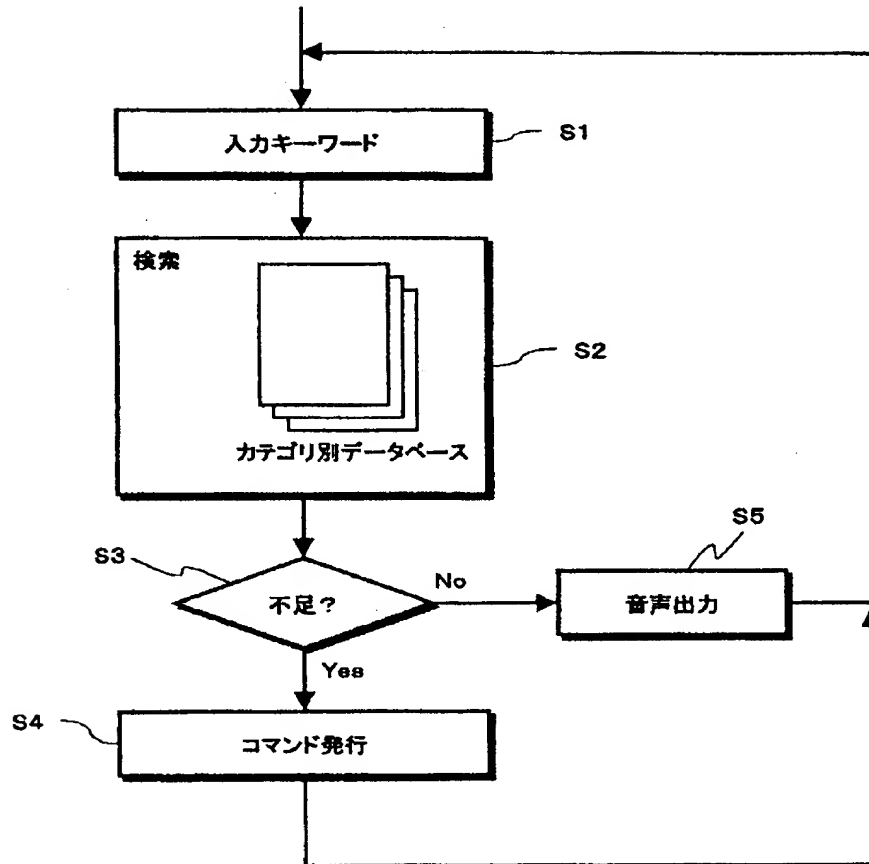
【図10】



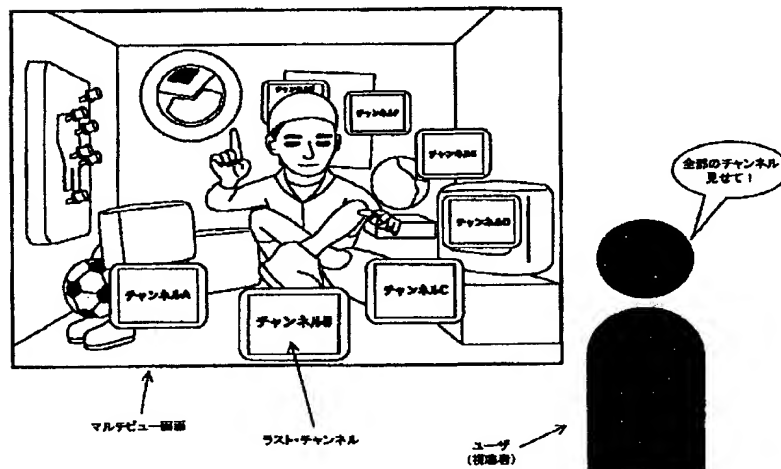
【図11】



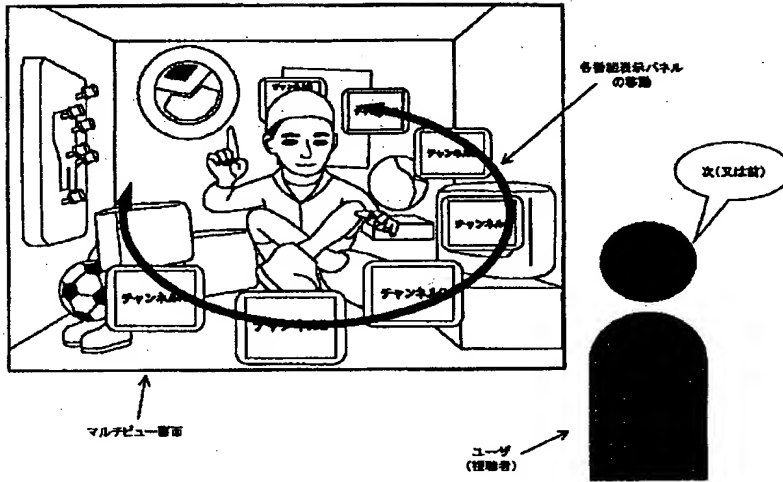
【図12】



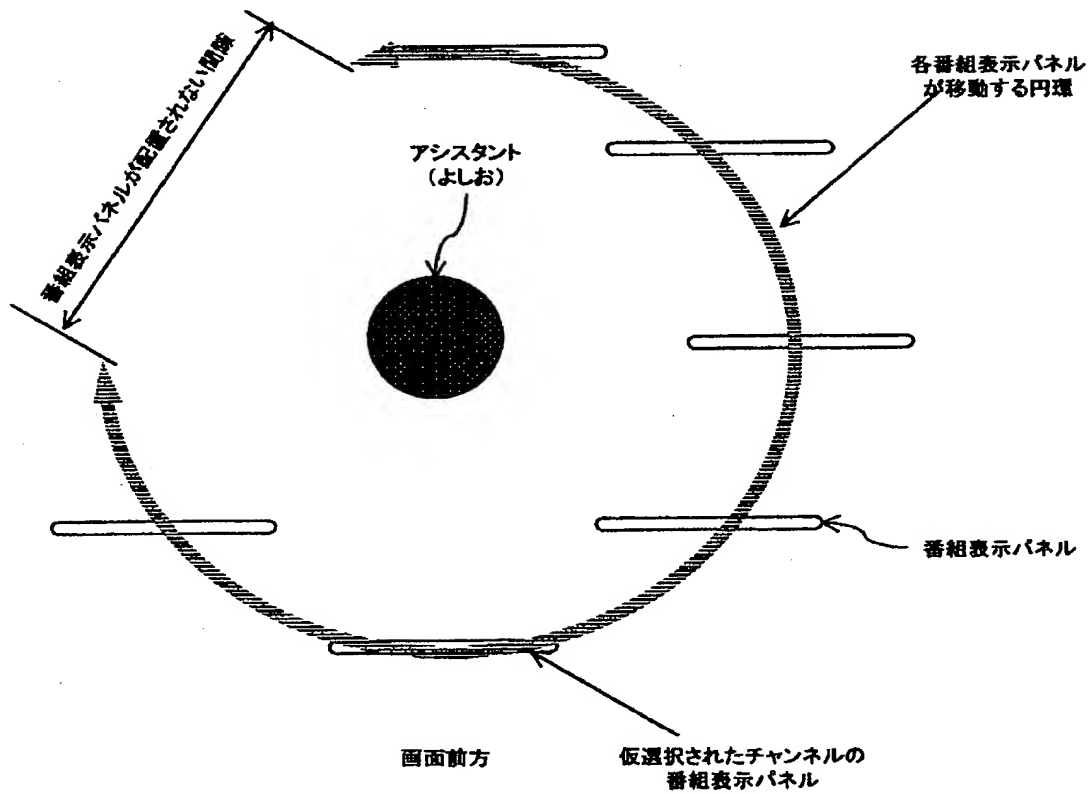
【図13】



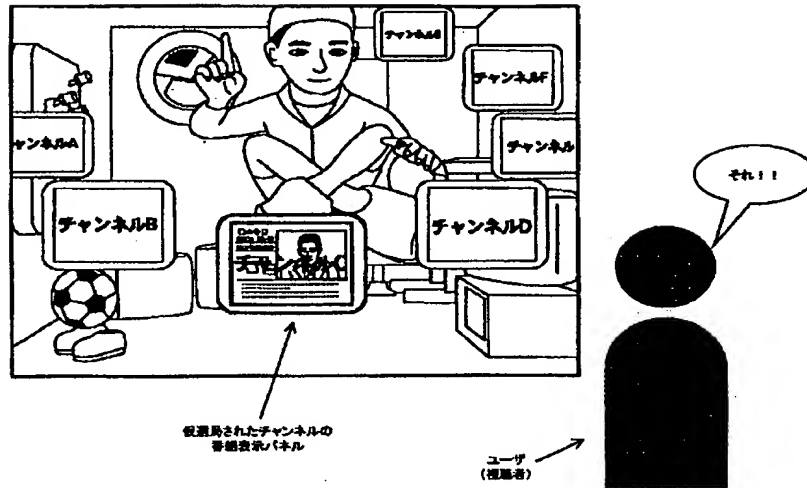
【図14】



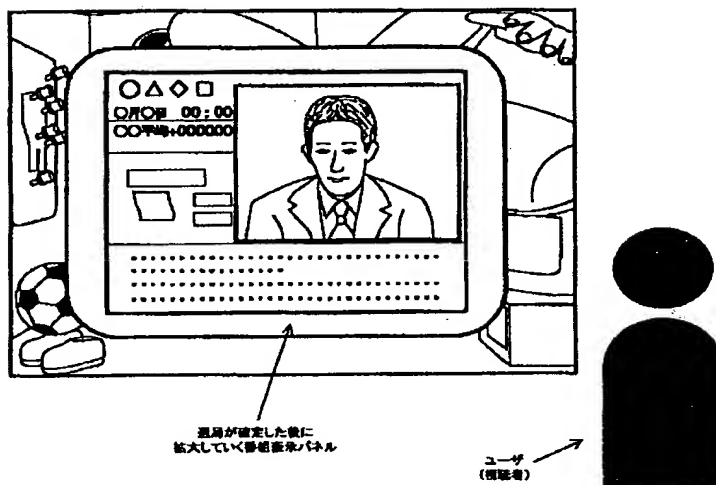
【図15】



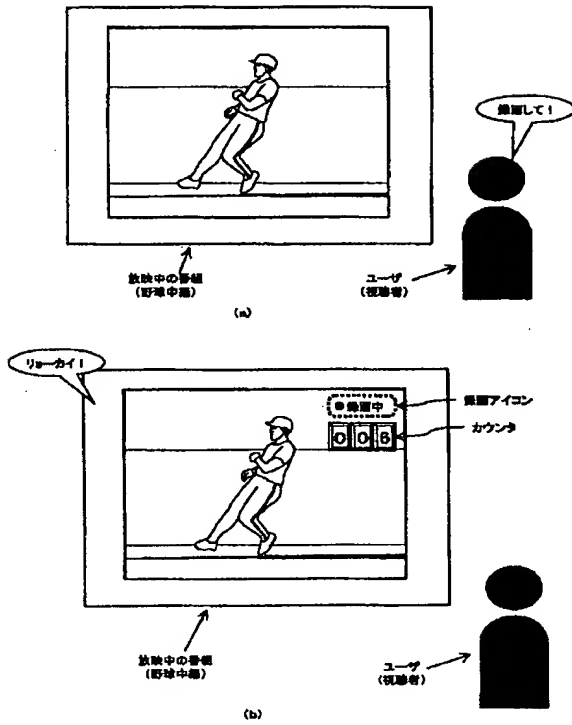
【図16】



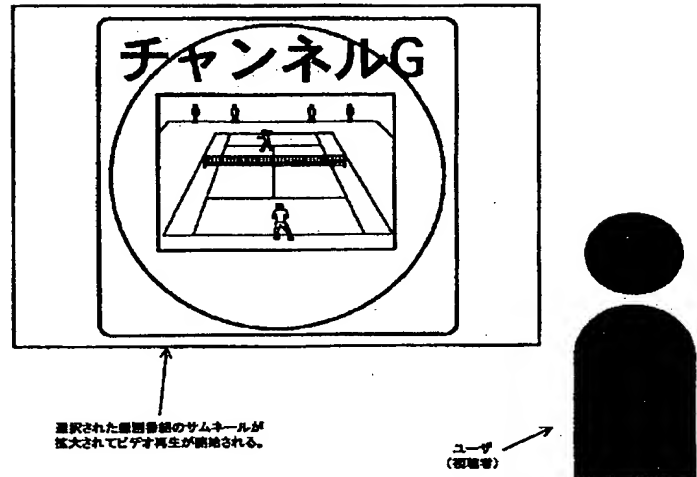
【図17】



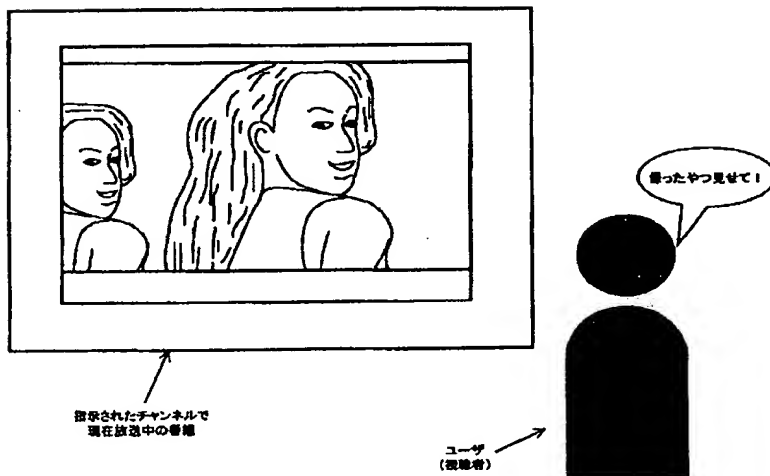
【図18】



【図22】

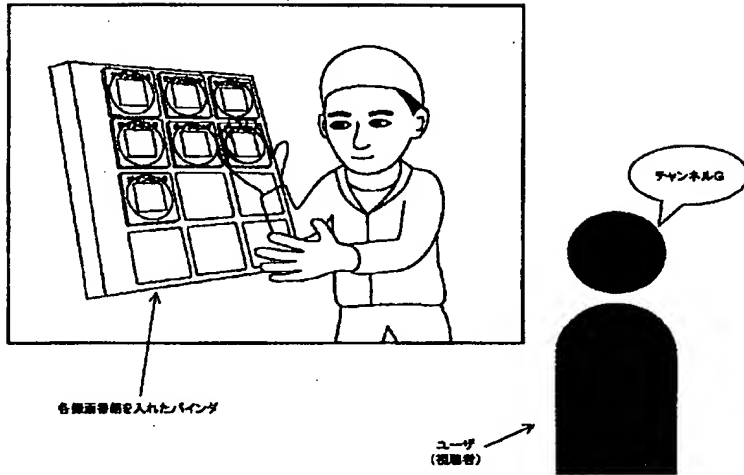


【図19】

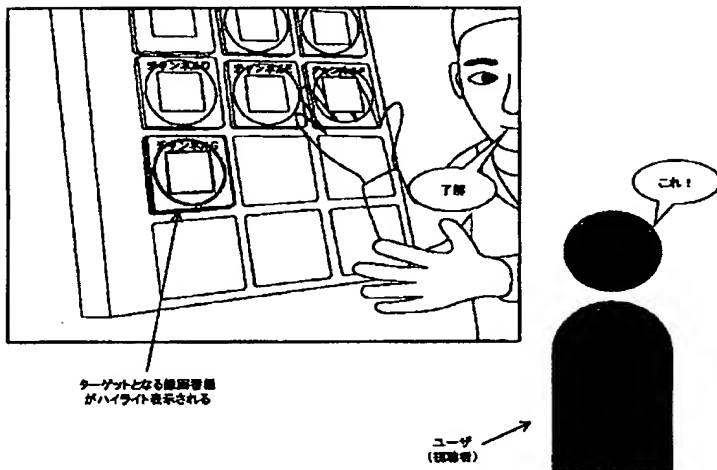




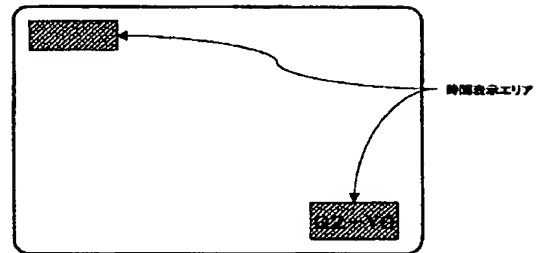
【図20】



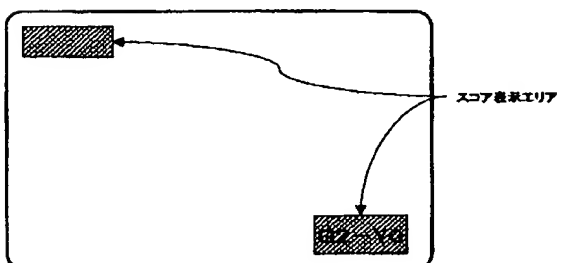
【図21】



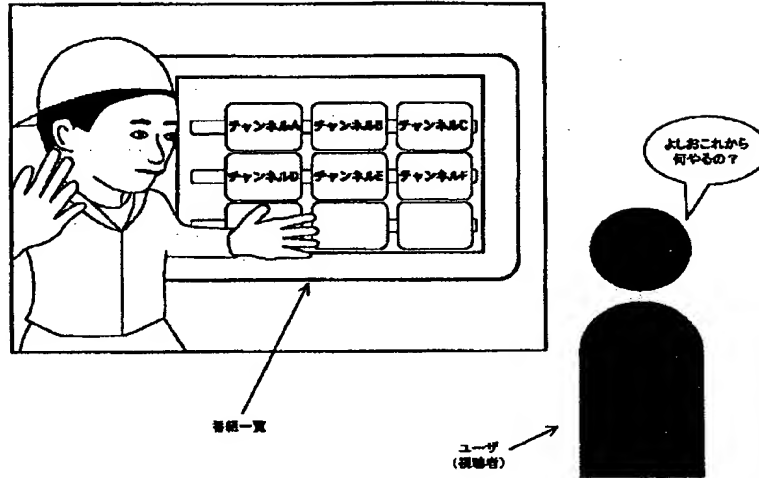
【図39】



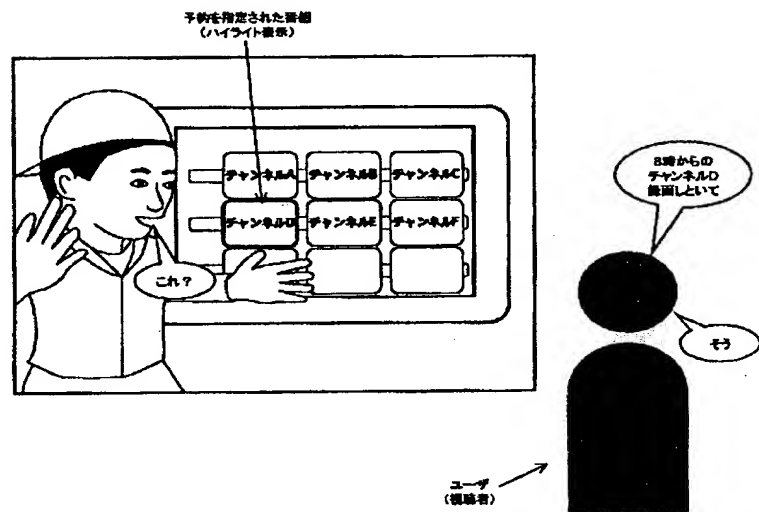
【図37】



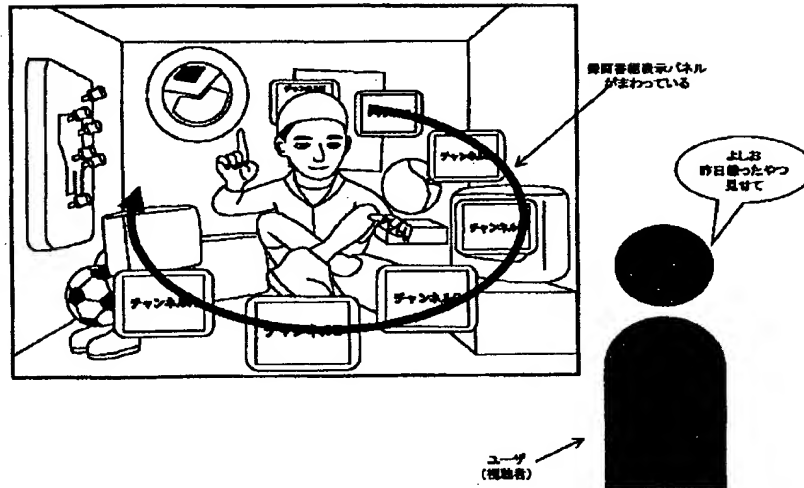
【図23】



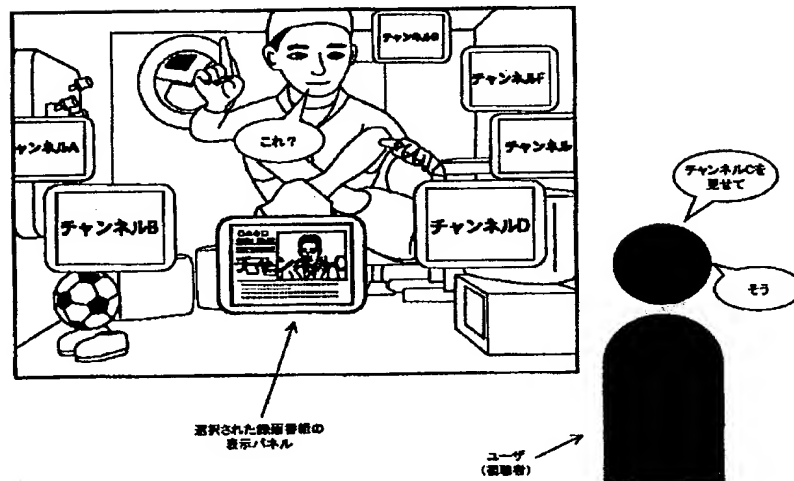
【図24】



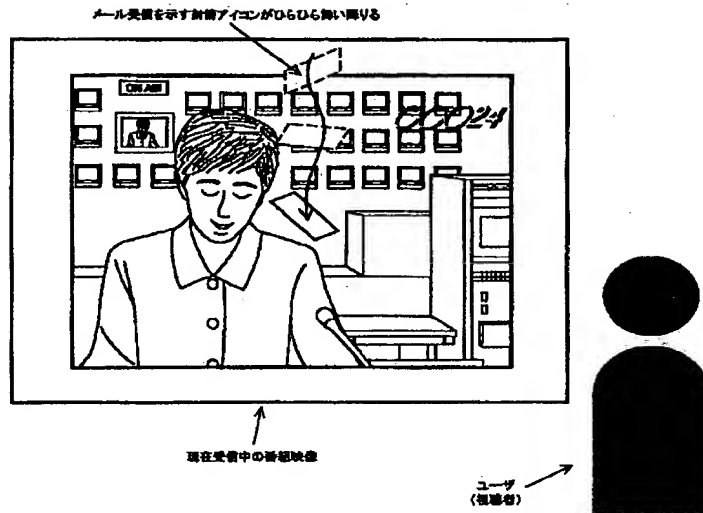
【図25】



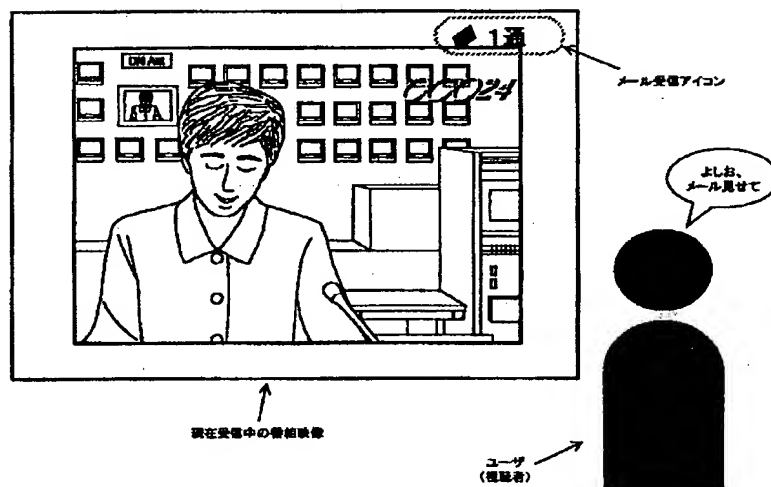
【図26】



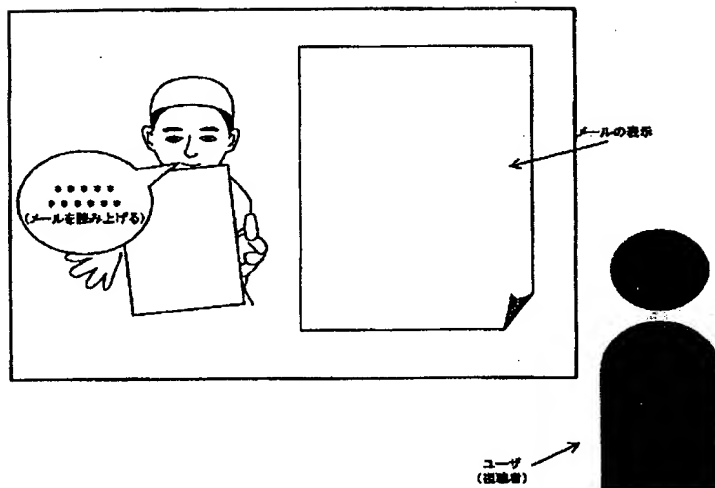
【図27】



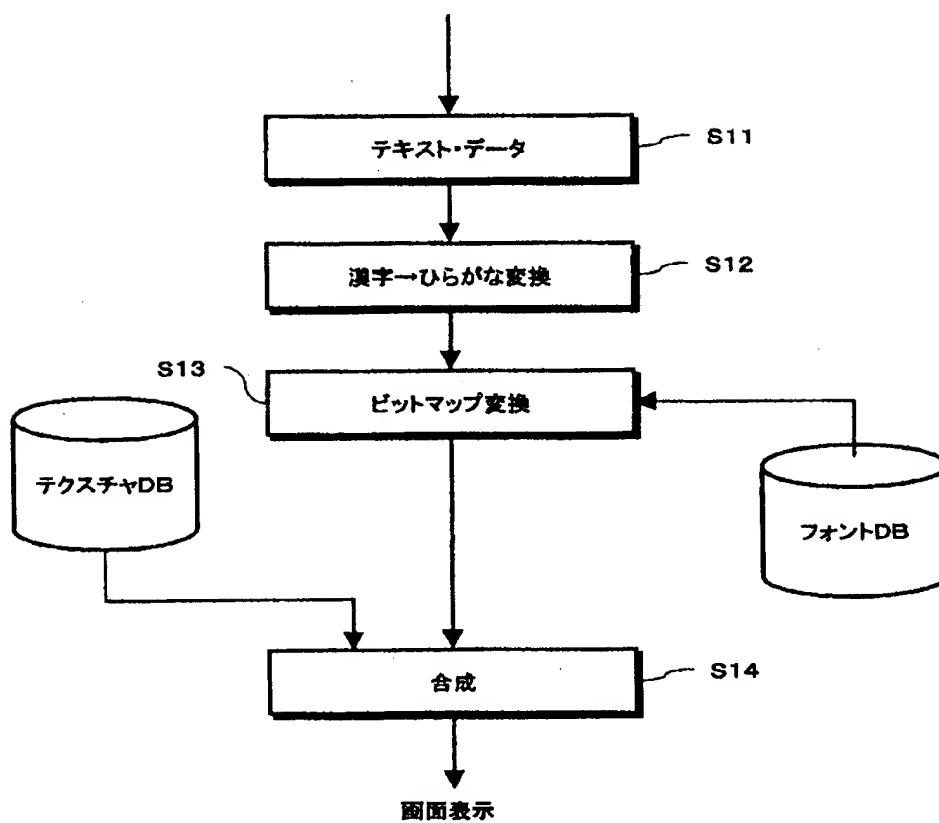
【図28】



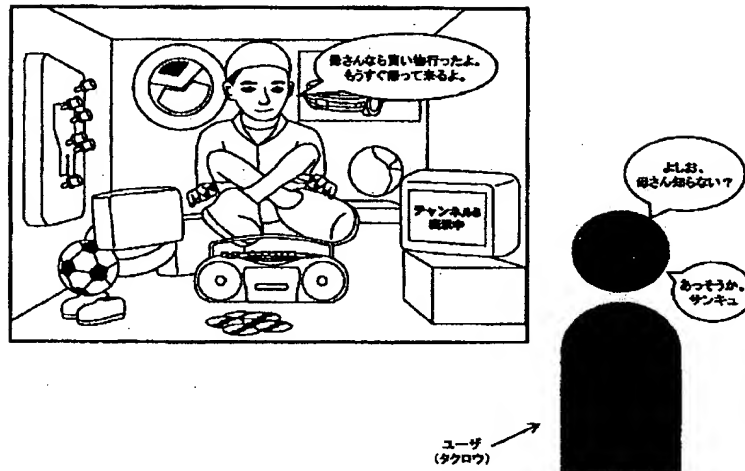
【図29】



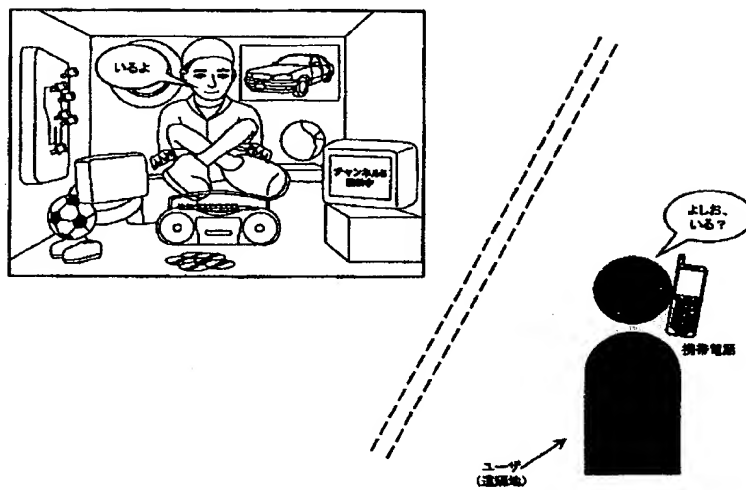
【図30】



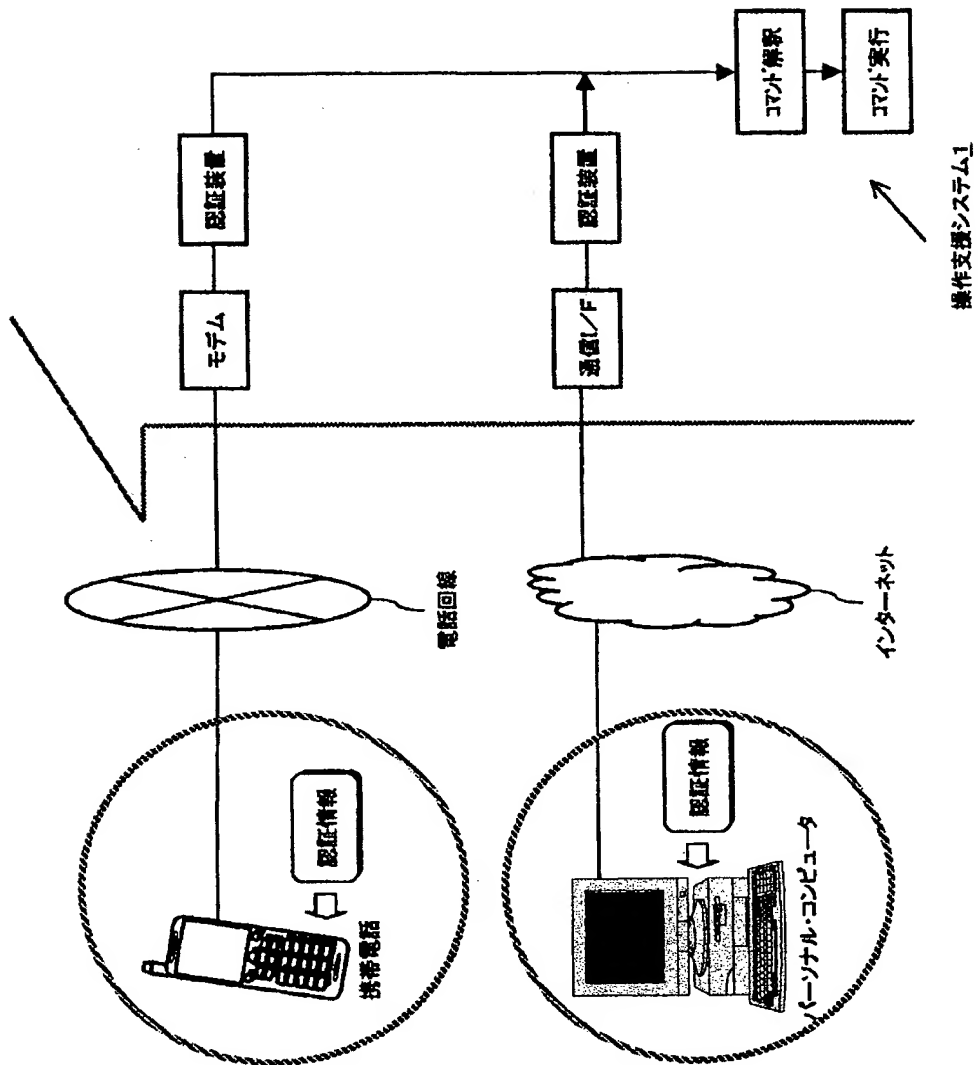
【図31】



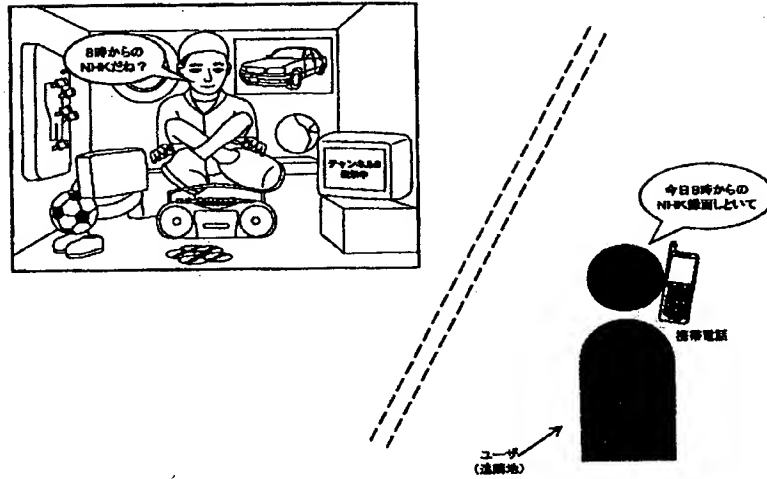
【図33】



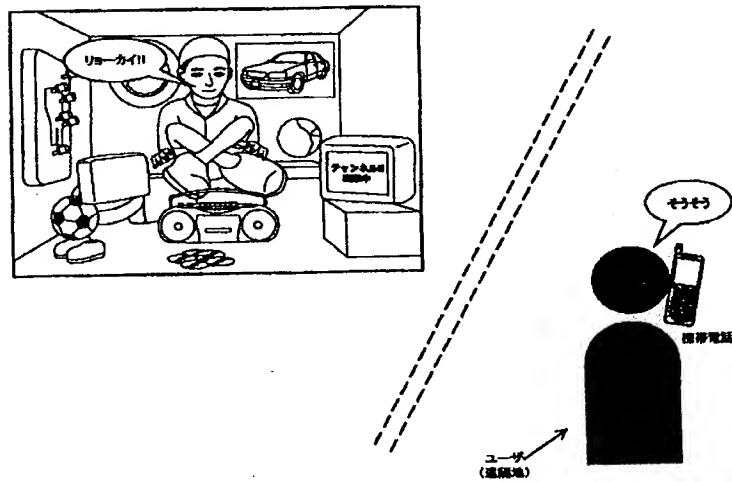
【図32】



【図34】

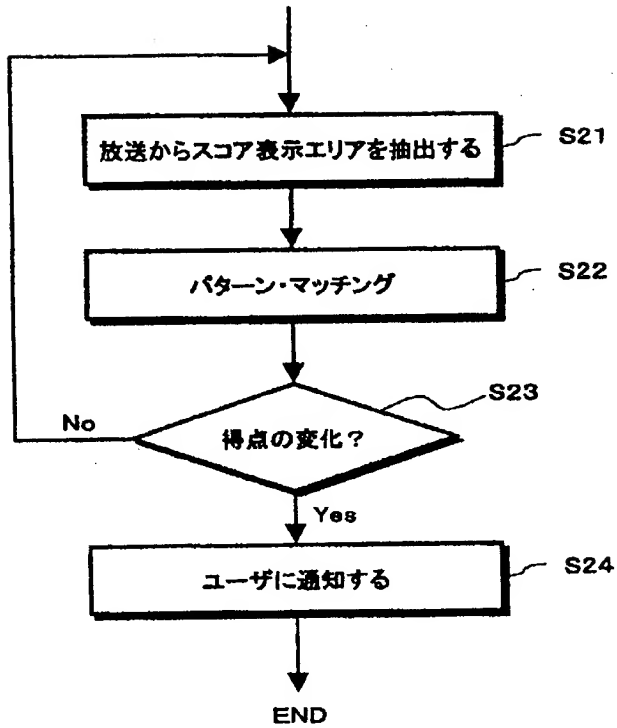


【図35】

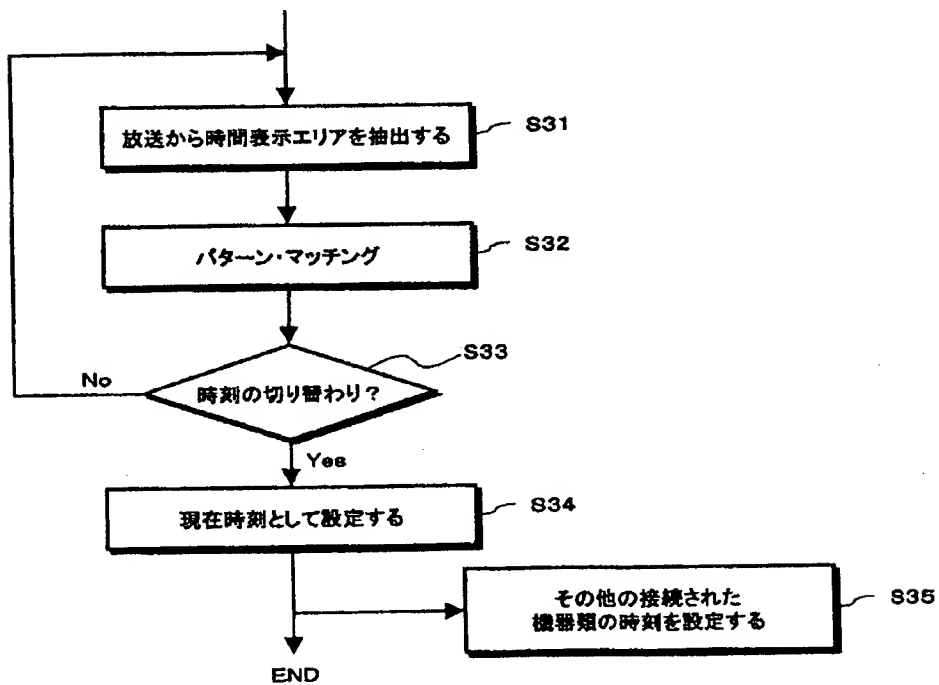




【図36】



【図38】



## フロントページの続き

(51)Int. Cl. <sup>7</sup>	識別記号	F I	ターマコード(参考)		
H O 4 N	5/00	H O 4 N	5/445	Z	5 K O 4 8
	5/44	H O 4 Q	9/00	3 2 1 E	
	5/445			3 3 1 A	
	7/025	G 1 0 L	3/00	5 5 1 G	
	7/03			5 5 1 A	
	7/035			5 7 1 T	
H O 4 Q	9/00	H O 4 N	7/08	A	
	3 2 1				
	3 3 1				

F ターム(参考) 5C025 BA28 CA02 CA09 CA19 CB05  
 CB09 DA08  
 5C056 AA10 BA01 DA08 EA09  
 5C063 AB07 AC01 DA03 EB03 EB33  
 EB37 EB40  
 5D015 KK01 KK02 LL10  
 5E501 AA02 AA19 AB03 BA05 CA02  
 CA08 CB15 EA21 FA45  
 5K048 BA03 DB05 DC07 EB02 EB14  
 FB10 HA04 HA06